MUSTERERKENNUNG VON STOCHASTISCHEN SIGNALEN BEHANDELT AM BEISPIEL VON BIOELEKTRISCHEN UND GEOPHYSIKALISCHEN AUFZEICHNUNGEN

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Georg-August-Universität zu Göttingen

vorgelegt von
Hans Michael Praetorius
aus Goslar

Göttingen 1976

Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41534923_0006

MUSTERERKENNUNG VON STOCHASTISCHEN SIGNALEN BEHANDELT AM BEISPIEL VON BIOELEKTRISCHEN UND GEOPHYSIKALISCHEN AUFZEICHNUNGEN

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Georg-August-Universität zu Göttingen

vorgelegt von
Hans Michael Praetorius
aus Goslar

Göttingen 1976

D 7

Referent: Prof. M. Siebert

Korreferent: Prof. M.R. Schroeder

Korreferent: Prof. O.D. Creutzfeldt

Tag der mündlichen Prüfung: 1.7.1976

INHALTSVERZEICHNIS

Seite

1. Einleitung 1

1.1 Stand der automatischen EEG-Analyse 3

1.2 Zweck und Ziel der Untersuchungen 5

2. Verfahren zur Merkmalsgewinnung 10

2.1 Methoden der linearen Prädiktion 11

2.1.1 Berechnung des Prädiktor-Polynoms 14

2.1.2 Interpretation der quadratischen Spektren 27

2.1.3 Bestimmung der Ordnung des Prädiktor-Polynoms 29

2.2 Erkennung von Nichtstationaritäten 32

2.2.1 Erkennung von Segmentgrenzen 33

2.2.2 Erkennung von Transienten 41

2.2.3 Segmentation und Transientenerkennung für mehrere Kanäle 45

2.2.4 Güte der Merkmale 47

3. Verfahren zur Klassifizierung 51

3.1 Automatische Klassifikation von Segmenten 53

3.1.1 Die Methode von Ward 54

3.1.2 Klassifikationsbeispiele 56

3.2 Klassifikation nach vorgegebenen Gesichtspunkten 60

3.2.1 Grundlagen der Diskriminanzanalyse 63

3.2.2 Klassifikation 66

3.3 Wahl der Merkmale 80

4. Programme 87

4.1 Datenhaltungssystem 88

Seite

4.2 Programme zur Merkmalsgewinnung 91

4.3 Programme zur Klassifikation 93

5. Zusammenfassung 95

Literaturverzeichnis 97

1. Einleitung

Im Jahre 1924 gelang es Hans Berger (1929) zum erstenmal, die rhytmische Gehirnaktivität, die später den Namen Elektroencephalogramm (EEG) erhielt, von der Kopfhaut abzuleiten. Heute werden EEGs routinemäßig für zahlreiche klinische Fragestellungen benutzt, wie zur Lokalisation von Gehirntumoren, als Diagnosehilfe bei Epilepsien oder zur Schlafüberwachung. Obwohl EEGs schwierig zu interpretieren sind, haben sie große klinische Bedeutung erlangt, besonders da sie ohne jeglichen operativen Eingriff gemessen werden können.

Leider ist es bis heute nicht gelungen, ein allgemeines Modell der EEG-Entstehung anzugeben, das quantitative Aussagen zuläßt. Qualitativ läßt sich aus dem Vergleich von Messungen an der Kopfhaut und im Cortex folgern, daß die Potentialschwankungen des EEGs als summierte postsynaptische Potentiale von cortikalen Neuronen gedeutet werden können (Creutzfeldt et al. 1964). Die Amplituden dieser Potentialschwankungen auf der Kopfhaut betragen zwischen 10 und 200 µV, sie können jedoch bis auf 1 mV während epileptischer Anfälle ansteigen. Die interessierenden Frequenzen liegen etwa in einem Bereich von 1 Hz bis 35 Hz, wobei jedoch für klinische Fragestellungen ein Bereich zwischen 1 Hz und 25 Hz vollkommen ausreicht. Um z.B. die Lage von Tumoren feststellen zu können, werden mehrere EEGs gleichzeitig von verschiedenen

Punkten der Kopfhaut abgeleitet und miteinander verglichen; die Zahl der Ableitungen variiert i.a. zwischen 2 und 16. EEGs können "unipolar" oder "bipolar" gemessen werden. Bei bipolaren Ableitungen wird die Potentialdifferenz zwischen zwei Elektroden auf der Kopfhaut gemessen, bei unipolaren Ableitungen dient eine feste Ohrelektrode als Referenz. Ein typisches Erwachsenen-EEG mit 6 unipolaren Ableitungen ist in Abb. 1 zu sehen.

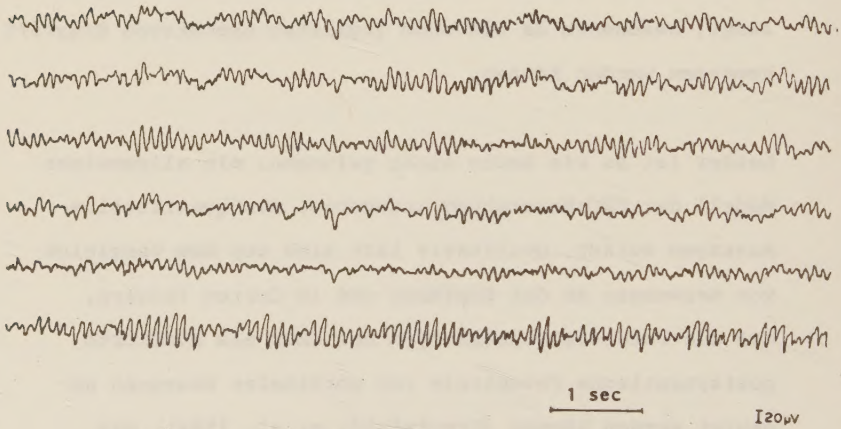

Abb. 1: Typisches EEG eines Erwachsenen mit sechs unipolaren Ableitungen

Der größte Teil aller EEG-Aufnahmen wird auf Papier ausgeschrieben und von speziell geschulten Ärzten analysiert, die als einziges Hilfsmittel eine Schablone benutzen, um Frequenzen messen zu können. Bei einem üblichen Papiervorschub von 3 cm/sec sind Wellen mit einer Frequenz von 30 Hz gerade noch 1 mm lang; der Papierschrieb einer typischen halbstündigen Registrierung ist länger als 50 m.

1.1 Stand der automatischen EEG-Analyse

Automatische Verfahren der EEG-Analyse sollen den Arzt von der Routineauswertung dieses immensen Materials möglichst befreien und ihm Hinweise für eine Diagnose liefern. Um die Analyse zu automatisieren, können wir auf zwei Arten vorgehen: entweder lassen wir uns von der Analyse des Arztes leiten und versuchen, sie weitgehend nachzuahmen, oder wir untersuchen das vorliegende Material mit bekannten Verfahren und versuchen erst anschließend, klinische Relevanz herzustellen. Beim ersten Ansatz ist es besonders schwierig, das intuitive Vorgehen des Arztes zu objektivieren und quantitativ zu formulieren, der zweite bleibt i.a. auf der Stufe stehen, daß mathematische Transformationen mit dem EEG durchgeführt werden und anschließend der Arzt diese Transformationen, deren Grundlagen ihm ein Geheimnis sind, auf ihre klinische Nützlichkeit hin beurteilen soll. Verfahren der Mustererkennung sind geeignet, die angesprochenen Probleme zu lösen.

Ein allgemeines System zur Mustererkennung ist in Abb. 2 dargestellt: das gemessene Signal wird durch relevante Merkmale charakterisiert, deren Zuordnung zu bestimmten Klassen bei EEGs die Umsetzung von mathematischen Merkmalen in klinisch verwertbare Aussagen bedeutet.

Abb. 2: Allgemeines System zur Mustererkennung

Zur Merkmalsgewinnung werden EEGs häufig im Zeitbereich beschrieben, wobei Symmetrie, Größe und Steigungen einzelner Wellen sowie deren Dauer durch Nulldurchgänge bestimmt werden können. Diese Methoden beruhen mehr oder weniger auf anschaulichen graphischen Überlegungen. Das EEG wird als Handschrift angesehen, deren einzelne Buchstaben - Grapho-Elemente - gedeutet werden müssen. Die zeitliche Struktur des EEGs bleibt vollständig erhalten, was allerdings mit einer nur geringen Datenreduktion erkauft wird.

Ein anderer Ansatz geht davon aus, daß sich EEGs als stochastische Signale deuten lassen, deren relevante Information durch quadratische Spektren (Power-Spektren) beschrieben werden kann. Die zeitliche Struktur wird kaum berücksichtigt, da die Spektren immer aus Segmenten gleicher zeitlicher Dauer berechnet werden.

Eine ausführliche Beschreibung dieser Methoden zur Merkmalsgewinnung wird von Matoušek (1973) gegeben.

Zur Charakterisierung der Spektren werden seit wenigen Jahren parametrische Verfahren benutzt, die dem EEG ein Modell zugrunde legen und die Modellparameter schätzen. Als Modelle werden AR- ("autoregressive") oder ARMA- ("autoregressive-moving average") Prozesse verwandt. ARMA-Modelle, wie sie Zetterberg (1973) benutzt, sind aufwendig und nur auf Großrechnern zu bestimmen; eine Echtzeitanalyse von EEGs ist mit ihnen nicht möglich. Die Bestimmung von

AR-Modellen ist dagegen sehr einfach und kann in Echtzeit erfolgen.

Nachdem das EEG durch den einen oder anderen Merkmalssatz beschrieben ist, ist eine Weiterverarbeitung nicht nur wünschenswert sondern unumgänglich, um der mathematischen Information klinisch diagnostische Bedeutung zuzuordnen. Die Arbeiten auf diesem Gebiet der EEG-Analyse sind nicht sehr zahlreich. Flik (1973) benutzt Grapho-Elemente zur Erkennung und Klassifizierung von diagnostisch wichtigen "spike-and-wave" Komplexen. Dubes (1970) benutzt ebenfalls Merkmale im Zeitbereich und untersucht systematisch verschiedene Klassifikationsmöglichkeiten. Gute Ergebnisse liefert die Bestimmung von Schlafstadien, da diese klinisch am besten definiert sind. Zur Klassifikation werden i.a. Merkmale aus dem Frequenzbereich wie vorherrschende Frequenz und Werte des Spektrums selbst herangezogen. Eine Übersicht dieser Methoden gibt Dubes (1970). Am weitesten fortgeschritten ist eine Methode (Matoušek et al. 1975), die zur Klassifikation Quotienten aus verschiedenen Frequenzbändern des Spektrums benutzt.

Alle Methoden müssen relativ viele Merkmale gewinnen und aus diesen die zur Klassifizierung wichtigsten, z.B. durch eine Faktorenanalyse bestimmen.

1.2 Zweck und Ziel der Untersuchungen

Ein System zur automatischen EEG-Analyse muß die für den

Arzt wichtige Information gewinnen und sie mathematisch so beschreiben, daß eine große Datenreduktion erreicht wird und die Analyse möglichst in Echtzeit durchgeführt werden kann. Es wäre wünschenswert, das Vorgehen des Arztes zu simulieren und als Diagnosehilfe dem Arzt verständliche Hinweise zu geben.

Nach Cobb (1963) bestimmt der Arzt aus dem EEG Frequenz und Amplitude der rhytmischen Komponenten, sowie deren Lage und Verteilung. Zusätzlich untersucht er Lage, Häufigkeit und Form von etwaigen Transienten - plötzlichen kurzzeitigen Veränderungen im EEG.

Um diesen Anforderungen gerecht zu werden, müssen wir das EEG adaptiv in möglichst homogene Segmente unterteilen sowie Transienten erkennen (Abb. 3). Die daraus zu gewinnenden Merkmale sollen Segmente und Transiente gut charakterisieren und für die weiteren Klassifizierungsaufgaben geeignet sein, insbesondere sollen die Merkmalssätze kurz sein und möglichst die gleiche Länge besitzen,

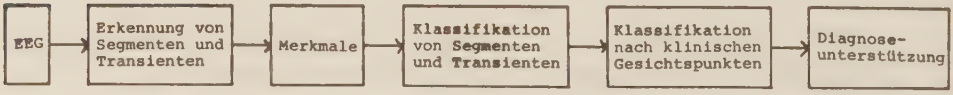

Abb. 3: System zur Mustererkennung bei EEGs

unabhängig von der Länge der einzelnen Segmente. Segmente und eventuell auftretende Transiente müssen anschließend klassifiziert werden, um ähnliche zusammenfassen zu können.

Mit diesen repräsentativen Segmenten und Transienten muß als letzter Schritt die Klassifizierung des gesamten EEGs nach klinischen Gesichtspunkten erfolgen, wie die Zuordnung zu verschiedenen Schlafstadien, Altersklassen oder Krankheiten.

Wie wir sehen werden, können mit einem AR-Modell Merkmale gewonnen werden, die diesen Anforderungen weitgehend gerecht werden.

Zur Segmentation liegt es nahe, ein AR-Modell mit variablen Koeffizienten anzuwenden, wie es von Kalman (1960) vorgeschlagen wurde. Mathieu (1974) untersucht an Hand von manuell segmentierten EEGs den zeitlichen Verlauf der Koeffizienten des Kalman-Filters. Er gibt jedoch kein Verfahren zur automatischen Segmentierung an. Duquesnoy (1975) benutzt ein geometrisches Verfahren, um mit den Koeffizienten des Kalman-Filters Grenzen automatisch festzulegen. Mit den benutzten Kriterien ist es zwar möglich, auf Änderungen im EEG zu schließen; jedoch ist es bis jetzt noch nicht gelungen, diese Änderungen quantitativ zu erfassen und daraus ein robustes automatisches Verfahren zu entwickeln. Darüberhinaus ist die Bedeutung des Kriteriums nicht klar. Isaksson (1975a,b) schließlich definiert einen Index, der die Variabilität der Koeffizienten des Kalman-Filters beschreibt, mit dessen Hilfe EEGs als "variabel" oder "nicht variabel" klassifiziert werden können. Die Variabilität bezieht sich allerdings

nur auf die Koeffizienten des Kalman-Filters, physikalisch wird sie nicht interpretiert. Auch mit diesem Verfahren ist bis jetzt noch kein automatisches Segmentierungsergebnis berichtet worden.

Alle drei Methoden haben den Nachteil, daß auf den heute zur Verfügung stehenden Rechnern keine Segmentierung in Echtzeit möglich ist. Prinzipielle Probleme bei der Benutzung von Kalman-Filtern, wie die Erkennung von kurzen Segmenten (kürzer als 5 sec) sollen hier nicht erörtert werden.

Wir werden ein AR-Modell benutzen, das stationäre Signale zu beschreiben gestattet. Mit zwei einfachen Kriterien können wir Änderungen im quadratischen Spektrum erkennen und kurzzeitige Veränderungen - Transienten - im Signal entdecken. In Kapitel 2 werden die dazu nötigen Verfahren vorgestellt und gezeigt, durch welche Merkmale Segmente und Transiente charakterisiert werden können.

In Kapitel 3 werden Möglichkeiten zur Klassifikation der gewonnenen Merkmale aufgezeigt. Der ersten Klassifikationsstufe, die keinerlei a-priori Information benötigt (Clusteranalyse), schließen sich einige Verfahren an, die es erlauben, die Merkmalssätze klinischen Gesichtspunkten zuzuordnen (Diskriminanzanalyse). Dazu stehen u.a. 340 EEGs aus vier verschiedenen Altersklassen zur Verfügung, von 4-, 6-, 10- und 12-jährigen Kindern.

Die Wahl der Merkmale hat mehr nach den erzielbaren Klassifikationsergebnissen zu erfolgen und nicht so sehr nach ihrer Interpretierbarkeit. In Kapitel 3.3 werden einige qualitative und quantitative Verfahren zur Merkmalsauswahl angegeben.

Kapitel 4 gibt schließlich eine Übersicht über die zum Teil sehr umfangreichen Programmsysteme.

Die Verfahren zur Merkmalsgewinnung und Klassifikation sind nicht nur auf bioelektrische Aufzeichnungen beschränkt, sondern sind prinzipiell bei stochastischen Signalen anwendbar, wenn nur die Modellvorstellung zutrifft, daß sich die Signale aus schwach stationären Segmenten zusammengesetzt denken lassen. Als weitere Anwendungsbeispiele werden geomagnetische Pulsationen untersucht. Mit dem Segmentationsverfahren ist es möglich, das Auftreten von plötzlichen Veränderungen in einer oder mehreren Komponenten zu entdecken. Die diskriminanzanalytischen Methoden zur globalen Klassifizierung können beim heutigen Stand der Forschung noch nicht sinnvoll eingesetzt werden, da das Interesse hauptsächlich darauf gerichtet ist, einzelne Pulsationen zu erkennen, zu beschreiben und ihre Beziehungen zu anderen geophysikalischen Phänomenen zu untersuchen.

2. Verfahren zur Merkmalsgewinnung

Verfahren zur automatischen EEG-Analyse müssen das Signal in Segmente unterteilen und kurzzeitige Veränderungen - Transiente - erkennen, wie etwa steile Wellen oder auch Artefakte. Die Signaleigenschaften sollen sich während eines Segmentes nicht oder nur möglichst wenig ändern, insbesondere wollen wir fordern, daß das erste und zweite Moment konstant bleiben. Die Verfahren zur Merkmalsgewinnung müssen folgende Forderungen erfüllen:

a) Aus den Segmenten, die (schwach) stationäre Signalteile beschreiben, muß die relevante Information gewonnen werden. Es bietet sich eine Beschreibung durch quadratische Spektren an.
b) Transiente müssen adäquat charakterisiert werden, wozu sich am besten ihre zeitliche Struktur eignet, da sie von diagnostischem Interesse sein kann.
c) Die gewonnenen Merkmalssätze müssen für eine weitergehende Verarbeitung geeignet sein, insbesondere sollten sie möglichst immer eine konstante Länge haben.
d) Die Datenreduktion soll schon bei der Merkmalsgewinnung groß sein; Segmente und Transiente sind durch möglichst wenige Werte zu beschreiben.
e) Die Merkmalsgewinnung muß schnell durchzuführen sein, damit das Verfahren auch praktische Anwendung erfährt.

Wie man diesen Forderungen gerecht werden kann, wird im Folgenden dargestellt.

2.1 Methoden der linearen Prädiktion

Ein lineares System mit der Übertragungsfunktion H(z) transformiert ein Eingangssignal u_n * zu einem Ausgangssignal s_n.

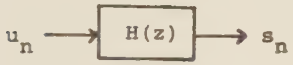

Abb. 4: Modell zur Erzeugung eines Signals s_n aus einem Eingangssignal u_n

Das gegenwärtig gebräuchlichste Modell berechnet s_n folgendermaßen:

$$s_n = - \sum_{k=1}^{p} a_k \, s_{n-k} + G \sum_{l=0}^{q} b_l \, u_{n-l} \quad \text{mit } b_0 = 1 \qquad (1)$$

Das Ausgangssignal s_n wird aus einer Linearkombination des schon bekannten Ausgangssignals und des gegenwärtigen und vergangenen Eingangssignals vorhergesagt; deshalb der Name lineare Prädiktion. Nach Anwendung der z-Transformation erhalten wir im Frequenzbereich:

$$S(z) = H(z) \cdot U(z) \quad \text{und}$$

$$H(z) = \frac{S(z)}{U(z)} = G \, \frac{1 + \sum_{l=1}^{q} b_l \, z^{-l}}{1 + \sum_{k=1}^{p} a_k \, z^{-k}} \qquad (2)$$

Die Wurzeln des Zählers sind die Nullstellen der Übertragungsfunktion, die des Nenners die Pole. Ein Nur-Pol-Modell

---

* Statt u(nT) (n = ganze Zahl, T = Abtastintervall) wird als Abkürzung u_n benutzt.

ist als AR-Modell ("autoregressive") bekannt; ein Modell, das nur durch Nullstellen charakterisiert wird, wird MA-Modell ("moving average") genannt. Das Modell mit der Übertragungsfunktion nach Gleichung (2) heißt daher ARMA-Modell.

Wir werden uns im Folgenden auf Nur-Pol- oder AR-Modelle beschränken, da sie gegenüber den ARMA-Modellen bestimmte Vorzüge haben (vgl. Makhoul 1975). Darüberhinaus können die Nullstellen der Übertragungsfunktion durch Pole dargestellt werden, wobei man allerdings eine größere Zahl von Polen berücksichtigen muß (Tretter, Steiglitz 1967). Ein System kann auch durch ein reines MA-Modell beschrieben werden, nur ist es i.a. sinnvoller, ein AR-Modell zu benutzen, da die Pole die Resonanzen des Systems beschreiben.

Bevor wir zur Berechnung der Koeffizienten des AR-Modells - des Prädiktor-Polynoms - übergehen, sollen einige Anwendungen dieses Modells in der Signalanalyse skizziert werden.

a) Anwendungen in der Geophysik

In der Ölexploration werden mit Hilfe von Explosionen künstliche Seismogramme erzeugt, aus denen sich Eigenschaften des Untergrundes ableiten lassen. Nach einem Vorschlag von Robinson (1967) kann ein Seismogramm als eine Summe von Impulsantworten verstanden werden. Als

Als Impuls dient eine Explosion, deren Reflexion an verschiedenen Untergrundstrukturen zu Impulsantworten (Wellenzügen) führt, deren Summe gerade das Seismogramm ergibt. Die Aufgabe besteht darin, das Seismogramm in einzelne Impulsantworten zu zerlegen. Durch diesen Prozeß der Dekonvolution erhält man eine Serie von unkorrelierten Impulsen, aus deren Ankunftszeit und Größe auf Reflexionshorizonte geschlossen werden kann.

Neuerdings richtet sich das Interesse mehr auf die Möglichkeit, Kurzzeitspektren zu berechnen, die wegen der hohen Streuung der einzelnen Spektralkomponenten mit den herkömmlichen Verfahren nur schwer zu bestimmen sind (Smylie et al. 1973; Drägert 1974).

b) Anwendungen in der Akustik

Der Stimmkanal ist in guter Näherung ein lineares Filter, das mit einem Nur-Pol-Modell beschrieben werden kann. Die Dekonvolution des Sprachsignals erlaubt es, auf einfache Weise die Grundfrequenz der Stimmbänder zu messen und die Form des Sprachkanals direkt zu bestimmen. Seit mehreren Jahren (Atal, Schroeder 1970) werden in immer größerem Umfang die Methoden der linearen Prädiktion benutzt. Eine Übersicht der Arbeiten geben z.B. Markel et al. (1973).

c) Anwendungen in der Neurophysiologie

Zur EEG-Analyse wurden bisher kaum AR-Modelle herangezogen.

Fenwick et al. (1971) versuchen evozierte Potentiale als Impulsantworten des Systems zu deuten und berechnen auch quadratische Spektren. Die Berechnung von Spektren wird noch von einigen anderen Autoren durchgeführt (Gersch 1970; Rappelsberger, Petsche 1975; Mathieu et al. 1976). Da mit einem AR-Modell aus den bekannten Werten des Signals der zukünftige Verlauf vorhergesagt werden kann, können unerwartete Veränderungen und speziell Transiente (Lopes da Silva et al. 1975) erkannt werden. Es ist bis jetzt nicht möglich, ein AR-Modell physikalisch oder neurophysiologisch zu erhärten, sodaß dieser Ansatz nur gerechtfertigt werden kann, weil er auf anderen Gebieten der Zeitreihenanalyse gute Erfolge gezeigt hat.

2.1.1 Berechnung des Prädiktor-Polynoms

Bei einem Nur-Pol-Modell gehen wir von Gleichung (1) aus, wobei alle Koeffizienten b_l, $l \neq 0$ identisch Null gesetzt werden:

$$s_n = - \sum_{k=1}^{p} a_k \, s_{n-k} + G \, u_n \qquad (3)$$

Die Übertragungsfunktion reduziert sich auf:

$$H(z) = \frac{G}{1 + \sum_{k=1}^{p} a_k \, z^{-k}} \qquad (4)$$

Der tatsächliche Wert des Signals zur Zeit nT, s_n, kann bis auf einen Fehler aus der Summe der vorhergehenden Werte vorhergesagt werden:

$$\hat{s}_n = - \sum_{k=1}^{p} a_k \, s_{n-k}$$

Der Prädiktionsfehler zur Zeit nT, ε_n, wird definiert als Differenz zwischen dem tatsächlichen Wert, s_n, und dem vorhergesagten, $\hat{s}_n$:

$$\varepsilon_n = s_n - \hat{s}_n = \sum_{k=0}^{p} a_k \, s_{n-k} \quad \text{mit } a_0 = 1 \tag{5}$$

Der Prädiktionsfehler beschreibt den mit einem AR-Modell nicht vorhersagbaren Anteil des Signals, er kann daher benutzt werden, um unerwartete Änderungen des Signals zu erkennen.

Um die Koeffizienten a_k - das Prädiktor-Polynom - bestimmen zu können, wird nach der Methode der kleinsten Quadrate der Erwartungswert des quadrierten Fehlers berechnet:

$$E\{\varepsilon_n^2\} = E\{(\sum_{k=0}^{p} a_k \, s_{n-k})^2\} \tag{6}$$

und minimiert:

$$\frac{\partial E}{\partial a_i} = 0; \quad 1 \leq i \leq p$$

Für i = 1 erhalten wir:

$$\frac{\partial E}{\partial a_1} = E\{2(\sum_{k=0}^{p} a_k \, s_{n-k}) \, \frac{\partial}{\partial a_1} (\sum_{k=0}^{p} a_k \, s_{n-k})\}$$

$$= 2\, E\{(\sum_{k=0}^{p} a_k\, s_{n-k})\ (s_{n-1})\}$$

$$= 2 \sum_{k=0}^{p} a_k\, E\{s_{n-k}\, s_{n-1}\}$$

Dieser Ausdruck wird gleich Null gesetzt:

$$\sum_{k=1}^{p} a_k\, E\{s_{n-k}\, s_{n-1}\} = -\, E\{s_n\, s_{n-1}\} \quad \text{für } i = 1$$

Allgemein erhalten wir folgendes Gleichungssystem:

$$\sum_{k=1}^{p} a_k\, E\{s_{n-k}\, s_{n-i}\} = -\, E\{s_n\, s_{n-i}\};\quad 1 \leq i \leq p \qquad (7)$$

Diese Gleichungen sind in der statistischen Literatur als Normalgleichungen bekannt. Den minimalen Erwartungswert des quadrierten Prädiktionsfehlers für ein gegebenes p, E_p, erhalten wir durch Kombination von (6) und (7):

$$E_p = E\{s_n^2\} + \sum_{k=1}^{p} a_k\, E\{s_n\, s_{n-k}\}$$

Wenn das Signal stationär ist, können die Erwartungswerte durch die Autokorrelationsfunktion ausgedrückt werden:

$$E\{s_{n-k}, s_{n-i}\} = R_{i-k}, \text{ wobei } s_n = 0 \text{ für } n \leq 0 \text{ und } n > N.$$

Daraus folgt:

$$\sum_{k=1}^{p} a_k\, R_{i-k} = -\, R_i;\quad 1 \leq i \leq p \qquad (8)$$

und

$$E_p = R_0 + \sum_{k=1}^{p} a_k R_k \qquad (9)$$

Eine elegante rekursive Lösung von (8) wurde zum erstenmal von Levinson (1949) angegeben.

Die Normalgleichungen erhalten wir aus Minimierung von (6). Eine andere Interpretation geht davon aus, daß die Erwartungswerte des quadrierten Fehlers für Vorwärts- und Rückwärtsprädiktion minimiert werden (Markel et al.1973). Für den Fehler, der bei Vorwärtsprädiktion entsteht, gelte:

$$\varepsilon^{+}_{m,n} = \sum_{i=0}^{m} a_{m,i}\, s_{n-i}; \quad a_{m,0} = 1$$

für den Fehler bei Rückwärtsprädiktion:

$$\varepsilon^{-}_{m,n} = \sum_{i=1}^{m+1} b_{m,i}\, s_{n-i}; \quad b_{m,m+1} = 1; \; 1 \leq m \leq p$$

p = Ordnung des Prädiktorpolynoms.

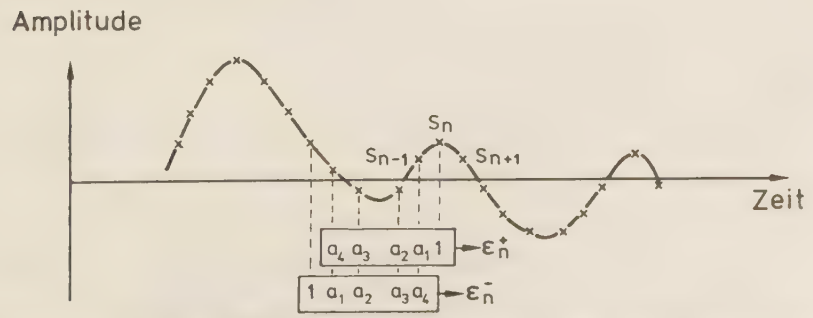

Abb. 5: Vor- und Rückwärtsprädiktion zur Bestimmung der Parcor-Koeffizienten. ε^{+}_{n} und ε^{-}_{n} sind die bei Vor- und Rückwärtsprädiktion entstehenden Prädiktionsfehler, die durch Faltung des Prädiktor-Polynoms (Länge:p=4) mit dem Signal berechnet werden.

Um das Gleichungssystem zu lösen, werden die Koeffizienten so gewählt, daß gilt (vgl. Abb. 5):

$$b_{m,k} = a_{m,m+1-k} \quad \text{für } k = 1, \ldots, m+1$$

Zur Bestimmung der $a_{m,i}$ und $b_{m,i}$ werden die Erwartungswerte der quadrierten Fehler simultan minimiert, was zu Rekursionsformeln führt:

$$\varepsilon^{+}_{m,n} = \varepsilon^{+}_{m-1,n} + k_m \, \varepsilon^{-}_{m-1,n} \quad \text{und}$$

$$\varepsilon_{m,n} = k_m \, \varepsilon^{+}_{m-1,n} + \varepsilon^{-}_{m-1,n-1}$$

Die dabei auftretenden Parcor-(partielle Korrelation) Koeffizienten sind definiert als:

$$k_m = - \frac{\sum_{n=-\infty}^{+\infty} \varepsilon^{+}_{m-1,n} \, \varepsilon^{-}_{m-1,n}}{\left(\sum_{n=-\infty}^{+\infty} \varepsilon^{+}_{m-1,n}{}^2 \cdot \sum_{k=-\infty}^{+\infty} \varepsilon^{-}_{m-1,n}{}^2 \right)^{1/2}} \qquad (10)$$

Die Parcor-Koeffizienten können als Korrelationskoeffizient zwischen s_n und s_{n+i} interpretiert werden, wenn $s_{n+1}, \ldots, s_{n+i-1}$ fest sind. Obwohl die Interpretation sehr verschieden ist von dem zuerst beschriebenen Ansatz, ist die Parcor-Methode äquivalent zu ihm. Die Parcor-Koeffizienten werden automatisch bei der Lösung der Normalgleichungen mit berechnet.

Da die Autokorrelationsmatrix positiv definit ist, müssen die Parcor-Koeffizienten immer kleiner als eins sein. Dies ist gleichzeitig eine notwendige und hinreichende Bedingung für die Stabilität der Übertragungsfunktion H(z) (Makhoul 1975).

Das von Burg (1967, 1968) angegebene Verfahren zur Bestimmung des Prädiktor-Polynoms geht ebenfalls wie die Parcor-Methode von einer Vorwärts- und Rückwärtsprädiktion aus. Der Fehler, der bei Vorwärtsprädiktion entsteht, ist definiert als (Smylie et al. 1973):

$$\varepsilon^{+}_{m,n} = \sum_{i=0}^{m} \gamma_{m,i}\, s_{n-i} \quad \text{mit } \gamma_{m,0} = 1$$

bei Rückwärtsprädiktion entsteht der Fehler (vgl. Abb. 6):

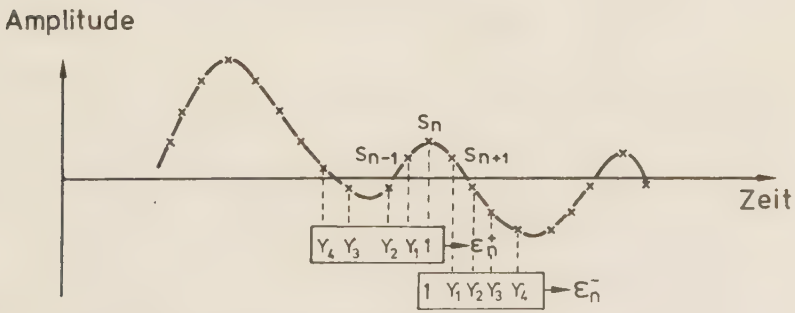

Abb. 6: Vor- und Rückwärtsprädiktion zur Bestimmung des "prediction-error-filters" nach Burg. Die Vor- und Rückwärtsprädiktionsfehler ε^{+}_{n} und ε^{-}_{n} entstehen durch Faltung des "prediction-error-filters" (Länge: p=4) mit dem Signal.

$$\varepsilon^{-}_{m,n} = \sum_{i=0}^{m} \gamma_{m,i}\, s_{n+i} \quad ; \quad 1 \leq m \leq p$$

Zur Bestimmung der $\gamma_{m,i}$ wird die Summe der Erwartungswerte der quadrierten Prädiktionsfehler minimiert, was ähnlich wie bei der Parcor-Methode zu rekursiven Bestimmungsgleichungen für den Prädiktionsfehler führt:

$$\varepsilon^{+}_{m,n} = \varepsilon^{+}_{m-1,n} + \gamma_{m,m}\, \varepsilon^{-}_{m-1,n-m} \quad \text{und}$$

$$\varepsilon^{-}_{m,n} = \varepsilon^{-}_{m-1,n} + \gamma_{m,m}\, \varepsilon^{-}_{m-1,n+m}$$

Die dabei auftretenden Koeffizienten $\gamma_{m,m}$ ("prediction-error-filter") sind:

$$\gamma_{m+1,m+1} = -\frac{2 \sum_{n=1}^{N-m+1} \varepsilon^{+}_{n+m+1,m} \cdot \varepsilon^{-}_{n,m}}{\sum_{n=1}^{N-m-1} \varepsilon^{+}_{n+m+1,m}{}^{2} + \sum_{n=1}^{N-m+1} \varepsilon^{-}_{n,m}{}^{2}}$$

N: Länge des Signals

$1 \leq m \leq p$

Im Vergleich zu den Parcor-Koeffizienten laufen die Summationsindices nicht von $-\infty$ bis $+\infty$, sondern benutzen nur die tatsächlich vorhandenen Werte des Signals, während bei der Parcor-Methode die Zeitreihe mit einem Fenster multipliziert wird, sodaß s_n für $n \leq 0$ und $n > N$ gleich Null wird. Die Koeffizienten $\gamma_{m,m}$ sind, wie man leicht aus ihrer Definition sehen kann, kleiner als eins. Bei der Parcor-Methode wird die Autokorrelationsfunktion explizit angegeben, bei der Methode nach Burg ist es nicht möglich, vielmehr berechnet sich die Autokorrelations-

funktion nach:

$$R_{m+1} = - \sum_{n=1}^{m+1} \gamma_{n,m+1} \, R_{m-n+1}$$

Ein Vergleich der Methoden nach Levinson und Burg ist in Abb. 7 zu sehen. Die Unterschiede sind für lange Zeitreihen bei der Autokorrelationsfunktion (links), dem Prädiktor-Polynom (Mitte) und den Spektren (rechts) unerheblich. Bei kurzen Zeitreihen können jedoch Unterschiede

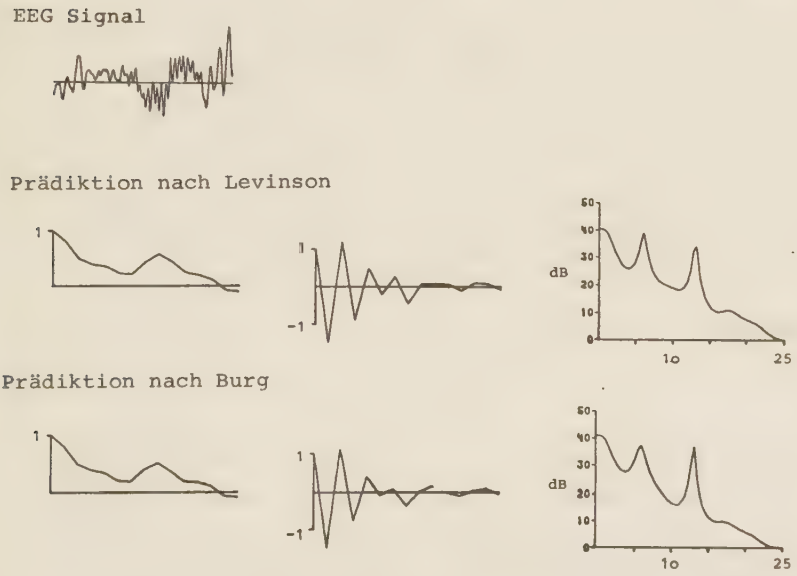

Abb. 7: Vergleich der Verfahren nach Burg und Levinson zur Berechnung eines Prädiktor-Polynoms. Länge des Signals: N=200, Länge des Polynoms: p=15. Links sind die Autokorrelationsfunktionen, in der Mitte die Prädiktor-Polynome und rechts die entsprechenden log. Spektren von 0-25 Hz aufgetragen. Die Unterschiede zwischen den Verfahren sind unerheblich.

auftreten. Die Spektren nach Burg zeigen schärfere Spitzen, deren Lage jedoch stärker streut. Wenn das Signal mit einem Fenster multipliziert wird (Verfahren nach Levinson), werden die Spitzen im Spektrum verschmiert, deren Lage dafür nicht so stark streut.

Daß die Lage der spektralen Spitzen bei kurzen Zeitreihen stark streuen kann, ist ein bekanntes Phänomen. Wenn wir eine volle Periode eines sinusförmigen Signals mit der Frequenz f analysieren, variiert die Frequenz im Spektrum je nach Phasenlage zwischen O,84·f und 1,12·f (Jackson 1967). Die Frequenzvariationen hängen von der Phasenlage und der Länge des untersuchten Signals ab; sie sind in Abb. 8 und 9 dargestellt. Untersucht wurde ein reines Sinussignal mit überlagertem weißen Rauschen (5% der Leistung des Sinussignals; das Verhältnis von Signal zu Rausch-

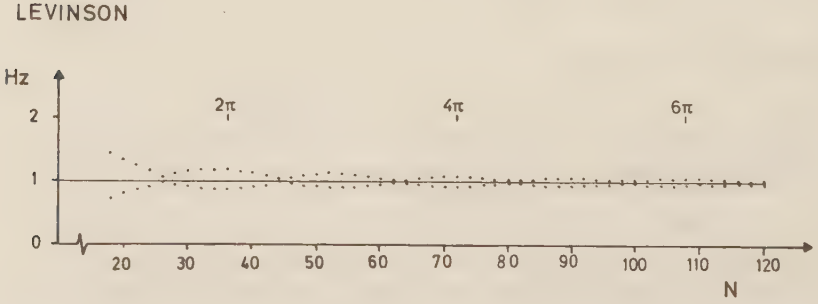

Abb. 8: Frequenzvariabilität nach dem Verfahren von Levinson. Sie hängt ab von der Phasenlage und Länge des Signals. Es sind für jedes N die kleinste und größte Abweichung für alle möglichen Phasenlagen O bis 2 π eines sinusförmigen Signals (1 Hz, 5% Rauschen) aufgetragen.

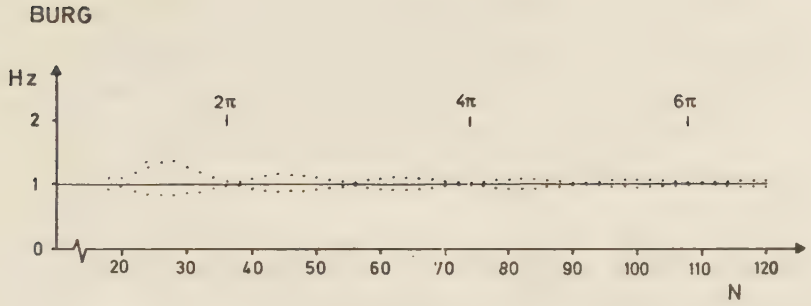

Abb. 9: Frequenzvariabilität nach dem Verfahren von Burg. Dasselbe Signal wie in Abb. 8.

leistung wirkt sich auf die Frequenzvariation nicht aus (Chen, Stegen 1974)). Die Methode nach Levinson führt zu maximalen Frequenzvariationen, wenn die Länge des Signals ein ganzzahliges Vielfaches von π beträgt (Abb. 8). Die Methode nach Burg besitzt für ganzzahlige Vielfache von π jeweils ein Minimum (Abb. 9). Daher ist es schwierig, die Spektren nach beiden Methoden für kurze Zeitreihen zu vergleichen.

Zur Erkennung von Segmentgrenzen bestimmen wir aus 100 Werten des Signals ein Prädiktorpolynom der Länge p=8. Die Signallänge entspricht bei einer maximalen Periode von 1.25 Hz (s.u.) etwa 3π. Aus den Abbildungen sehen wir, daß nach beiden Methoden Frequenzvariationen von etwa 10 bis 15% auftreten können. Die Unterschiede zwischen beiden Methoden sind bei diesen Werten also unerheblich. Die Zeit zur Berechnung des Prädiktor-Polynoms variiert allerdings sehr stark. Um aus 100 Werten ein Polynom der

Länge p=9 zu berechnen, werden im Mittel auf der UNIVAC 1108 mit der Methode nach Burg 0.041 sec benötigt, mit der Methode nach Levinson nur 0.013 sec. Wir benutzen deshalb die Methode nach Levinson zur Berechnung des Prädiktor-Polynoms.

Das quadratische Spektrum eines Systems mit der Übertragungsfunktion H(z) ist definiert als:

$$P(\omega) = H(e^{i\omega})\ H^{*}(e^{i\omega}) = |H(e^{i\omega})|^2$$

Mit der Übertragungsfunktion nach Gleichung (4) erhalten wir:

$$P(\omega) = \frac{G^2}{\left|1 + \sum_{k=1}^{p} a_k\ e^{-i\omega k}\right|^2} \qquad (11)$$

Unter der Annahme eines Gausschen Prozesses mit Mittelwert Null konnte Kromer (1969) zeigen, daß die Zahl der Freiheitsgrade für nach Gleichung (11) berechnete Spektren gleich N/p ist (N: Länge des Signals, p: Länge des Prädiktorpolynoms). Das Spektrum kann also berechnet werden, indem eine Konstante G^2 durch den quadrierten Absolutbetrag der Fouriertransformierten des Prädiktor-Polynoms dividiert wird. Wenn dessen Länge z.B. p=10 ist, wird man, um genug Werte zur Darstellung des Spektrums zu erhalten und die schnelle Fouriertransformation benutzen zu können, noch 54 oder 118 Nullen dem Polynom

hinzufügen. Abb. 10 zeigt das Spektrum eines 200 Werte langen EEG-Segmentes für verschiedene Längen des Prädiktor-Polynoms.

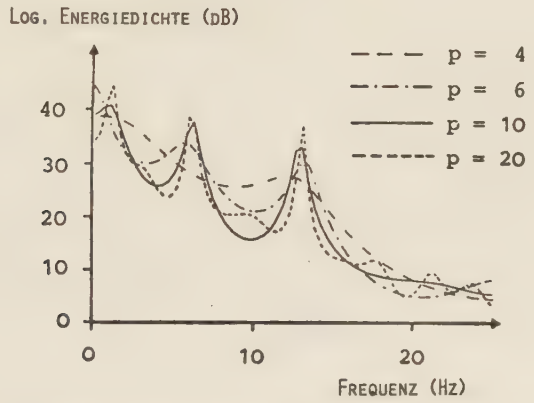

Abb. 10: Quadratische Spektren eines EEG-Segmentes für verschiedene Längen des Prädiktor-Polynoms. Länge des Segmentes: N=200; p=10 reicht zur Charakterisierung des Spektrums aus.

Das Prädiktor-Polynom wird auch inverses Filter genannt, da es jede Zeitreihe s_n in ein Signal mit einem flachen Spektrum, nämlich in ein weißes Rauschen transformiert (vgl. Gleichung 5). Für ein unendlich langes Prädiktor-Polynom ($p \to \infty$) erhalten wir mit Gleichung (11) exakt das inverse Spektrum der Zeitreihe (Markel et al. 1973). Für ein endliches p kann das Prädiktor-Polynom nur approximativ das inverse Spektrum der Zeitreihe charakterisieren; das Signal s_n wird in eine möglichst gute Realisierung des weißen Rauschens transformiert. Diese Realisierung besteht aus den nicht vorhersagbaren Anteilen, den Prädiktionsfehlern ε_n.

Abb. 11 soll diesen Sachverhalt verdeutlichen. Das Signal wird mit Hilfe des inversen Filters in zwei Anteile zerlegt: einen, der nur Rauschen repräsentiert (rechts), und einen, in dem kein Rauschanteil mehr vorhanden ist (links).

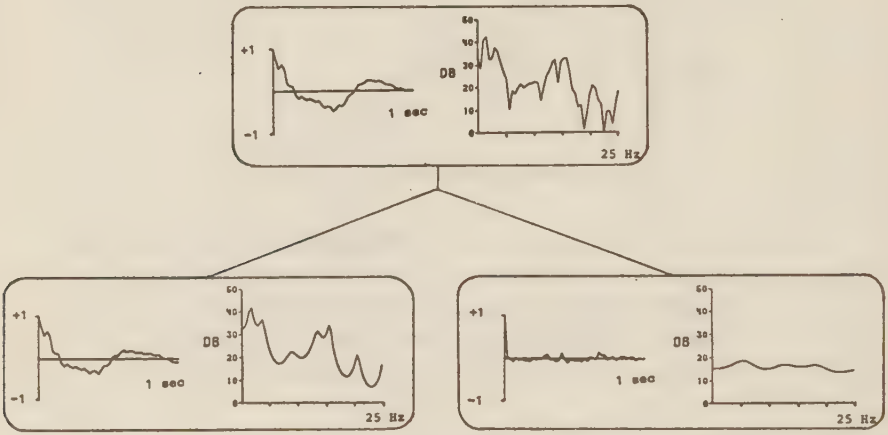

Abb. 11: Zerlegung eines Signals mit dem inversen Filter in einen Anteil mit und einen ohne Rauschen. Länge des EEGs: N=200, Länge des Prädiktor-Polynoms: p=20. Oben: Autokorrelationsfunktion und FFT-Spektrum des Signals. Unten, links: Autokorrelationsfunktion und Spektrum des Signalanteiles ohne Rauschen. Unten, rechts: Autokorrelationsfunktion und Spektrum des Rauschanteiles.

Wir sind jetzt in der Lage, den Faktor G in Gleichung (11) zu bestimmen. Ein Vergleich von (3) und (5) liefert:

$$s_n = - \sum_{k=1}^{p} a_k \, s_{n-k} + Gu_n \qquad \text{und}$$

$$s_n = - \sum_{k=1}^{p} a_k \, s_{n-k} + \varepsilon_n$$

Es ist sinnvoll zu fordern, daß die Gesamtleistung des Signals Gu_n gleich der Gesamtleistung des Prädiktionsfehlers sein muß, die durch Gleichung (9) gegeben ist. Da das Spektrum von u_n flach ist, wenn das Prädiktor-Polynom lang genug ist, kann G berechnet werden (Makhoul, Wolf 1972):

$$G^2 = E_p = R_o + \sum_{k=1}^{p} a_k \, R_k$$

2.1.2 Interpretation der quadratischen Spektren

Im letzten Kapitel wurde gezeigt, wie quadratische Spektren mit Hilfe des Prädiktor-Polynoms berechnet werden können; eine Interpretation der Spektren gibt Burg (1976,1968): Bei praktischen Problemen haben wir es immer mit endlichen Zeitreihen zu tun, wir können nur eine kleine Zahl der Koeffizienten der Autokorrelationsfunktion verläßlich schätzen. Sollen die übrigen Koeffizienten auf Null gesetzt werden oder sollen sie periodisch fortgesetzt werden, wie es in den üblichen Methoden der Zeitreihenanalyse geschieht? Beide Vorgehensweisen sind unbefriedigend, da sie dem Signal zusätzlich Information hinzufügen. Es scheint am sinnvollsten, die unbekannten

Koeffizienten so zu schätzen, daß keinerlei zusätzliche Information gebraucht wird, bzw. daß die Entropie des Prozesses maximal oder die Entropierate stationär wird.

Wir benutzen zur Bestimmung eines AR-Modells nur die ersten beiden Momente (vgl. Gleichung 8). Für diese Art von Prozessen, und zwar speziell für Gauß'sche Prozesse, kann die Entropie oder die Entropierate angegeben werden. Die Stationaritätsbedingung bedeutet, daß die partiellen Ableitungen der Entropierate nach den Autokorrelationskoeffizienten zu Null werden. Mit diesen Überlegungen ist es möglich, Gleichung (11) für die quadratischen Spektren abzuleiten (Smylie et al. 1973).

Die unbekannten Koeffizienten können nach Gleichung (8) für i > p bestimmt werden (Robinson 1967). Für die Berechnung des Prädiktor-Polynoms werden nur die ersten p Koeffizienten benötigt. Die Koeffizienten für größere Zeitverschiebungen werden so bestimmt, daß dem Prozeß keine zusätzliche Information hinzugefügt wird (Abb. 12).

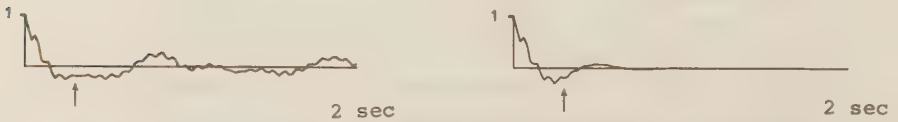

Abb. 12: Maximum Entropie Schätzung der Autokorrelationsfunktion. Links: Autokorrelationsfunktion des Originalsignals. Rechts: Die ersten 15 Werte der Autokorrelationsfunktion (Pfeil) werden zur Bestimmung des Prädiktor-Polynoms benutzt, mit dem die Koeffizienten für größere Zeitverschiebungen bestimmt werden.

2.1.3 Bestimmung der Ordnung des Prädiktor-Polynoms

Es wäre wünschenswert, die Ordnung des Prädiktor-Polynoms p auf Grund von physikalischen Gesichtspunkten abschätzen zu können. Da aber keine allgemeine Theorie der EEG-Entstehung existiert, ist dieses Vorgehen nicht möglich. Die Ordnung p kann jedoch durch Untersuchung des Erwartungswertes des minimalen quadratischen Fehlers E_p (Gleichung 9) geschätzt werden. Der normalisierte Fehler V_p sei definiert als:

$$V_p = \frac{E_p}{R_0} = 1 + \sum_{k=1}^{p} a_k r_k \quad \text{mit} \quad r_k = \frac{R_k}{R_0}$$

V_p ist auf den Wertebereich $0 \leq V_p \leq 1$ beschränkt (Robinson, Treitel 1969). Ferner kann gezeigt werden, daß V_p für wachsende Ordnung des Prädiktor-Polynoms immer kleiner werden muß (Levinson 1949):

$$0 \leq V_p \leq V_{p-1} \quad \text{mit} \quad V_0 = 1$$

$V_p = 1$ bedeutet, daß das Signal nicht vorhersagbar ist, also ein weißes Rauschen darstellt mit einem flachen Spektrum; für $V_p = 0$ ist das Signal exakt vorhersagbar. Der kleinste Wert von V_p, nämlich V_∞, läßt sich bestimmen (Makhoul, Wolf 1972):

$$V_{min} = V_\infty = \frac{e^{c_0}}{R_0}$$

wobei c_0 der erste Koeffizient des Cepstrums (inverse

Fouriertransformierte des log. Spektrums) ist. Wenn das Spektrum des Signals nur für diskrete Werte bekannt ist, läßt sich dieser Ausdruck umformen zu:

$$V_{min} = \frac{\left(\prod_{k=1}^{N} P_k\right)^{1/N}}{\frac{1}{N}\sum_{k=1}^{N} P_k} \quad ; \; P_k = \text{k-ter Wert des Spektrums}$$

Dieser Ausdruck ist nichts anderes als der Quotient von geometrischen zu arithmetrischen Mittel. In Abb. 13 sind die Werte für V_{min} als gestrichelte Linien eingezeichnet.

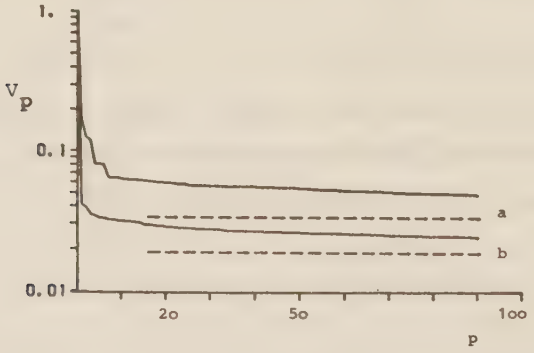

Abb. 13: Abhängigkeit des normierten Erwartungswertes des quadrierten Prädiktionsfehlers von der Länge des Prädiktor-Polynoms. V_{min} $(p \to \infty)$ ist gestrichelt eingezeichnet. Das zu b) gehörende EEG ist besser vorhersagbar, als das zu a) gehörende.

Aus dem Verlauf von V_p kann die Ordnung p bestimmt werden. Gewöhnlich wird angenommen, daß das Signal sich durch ein AR-Modell mit der wahren Ordnung p darstellen läßt;

das Problem liegt darin, p zu schätzen. Gersch (1970) gibt ein Verfahren an, daß p unter Berücksichtigung der relativen Änderung von V_p schätzt. Danach braucht p bei EEGs nicht größer als 10 zu sein. Akaike (1971) schlägt vor, p durch m zu schätzen, wenn folgendes Kriterium minimiert ist:

$$AIC(m) = \log V_m + 2\frac{m}{T} \qquad (12)$$

m: Ordnung des Prädiktor-Polynoms

T: Länge der Zeitreihe, aus der die Autokorrelationsfunktion bestimmt wird.

Nach Parzen (1974) kann ein Signal als AR-Prozeß für $p = \infty$ aufgefaßt werden, wobei auch Nullstellen in der Übertragungsfunktion auftreten können, die man durch Pole darstellen kann (es stehen für $p = \infty$ beliebig viele zur Verfügung). Um eine optimale Schätzung der Übertragungsfunktion zu bekommen, schlägt er vor, folgendes Kriterium zu minimieren:

$$CAT(m) = 1 - \frac{V_{min}}{V_m} + \frac{m}{T} \qquad (13)$$

Nach Akaikes Interpretation ist die Ordnung m eine Schätzung der wahren Ordnung p, während Parzen die Übertragungsfunktion eines AR-Modells für $p = \infty$ durch die Übertragungsfunktion eines AR-Modells der Ordnung m schätzt. Die nach diesen Verfahren bestimmten Ordnungen differieren nicht sehr stark; CAT liefert, wenn überhaupt,

Unterschiede auftreten, einen größeren Wert.

Tabelle 1:

	AIC	CAT
EEG, Spektrum mit 3 Spitzen	8	9
EEG, ausgeprägte 14 Hz Komponente	4	4
EEG, 7 - 8 Hz Aktivität	4	6
EEG, geringe Aktivität	6	6
EEG, sehr unregelmäßige Aktivität	7	12
EEG, normale Aktivität	6	6

Für EEGs scheint eine Ordnung von p=8 angemessen zu sein. Bei geomagnetischen Pulsationen erhalten wir i.a. etwas größere Ordnungen:

Tabelle 2:

	AIC	CAT
Pulsationen, geringe Aktivität	8	8
Pulsationen, starke Aktivität	8	8
Pulsationen, sehr unregelmäßige Aktivität	10	12

2.2 Erkennung von Nichtstationaritäten

Wir nehmen als zugrunde liegende Vorstellung an, daß sich EEGs als eine Folge von Segmenten unterschiedlicher Dauer beschreiben lassen, wobei die statistischen Parameter der einzelnen Segmente - insbesondere das quadratische Spektrum - nicht oder nur in geringem Umfang variieren. Starke Änderungen sollen zu Segmentgrenzen führen. Darüberhinaus können kurzzeitige Veränderungen - Transiente - den an sich

stationären Segmenten überlagert sein. Wir werden zeigen, wie es möglich ist, mit Hilfe eines AR-Modells, das nur für stationäre Signale definiert ist, Segmentgrenzen und Kurzzeitnichtstationaritäten zu entdecken. Dazu werden die Eigenschaften des Prädiktionsfehlers (Gleichung 5) untersucht, der für stationäre Signale sich als weißes Rauschen beschreiben läßt. Korrelationen im Prädiktionsfehler liefern ein Kriterium zur Erkennung von Segmentgrenzen; mit den Werten des Prädiktionsfehlers selbst können Transiente erkannt werden.

2.2.1 Erkennung von Segmentgrenzen

Segmentgrenzen sollen erkannt werden, wenn die Leistung des Signals sich plötzlich ändert; dieses Problem ist einfach zu lösen. Sie sollen aber auch erkannt werden, wenn bei gleichbleibender Leistung sich nur die Frequenzen ändern. Die der Grenzenerkennung zugrunde liegende Idee ist folgende: Wir stellen uns ein über den Anfang der Zeitreihe gelegtes Fenster der Länge N vor (Abb. 14 A). Aus den durch dieses Fenster gesehenen Werten wird eine Autokorrelationsfunktion berechnet, anschließend werden die Normalgleichungen (8) gelöst und das Prädiktor-Polynom bestimmt. Der Prädiktionsfehler hat, wenn die Ordnung des Prädiktor-Polynoms richtig gewählt ist, während der ersten N Werte ein flaches Spektrum, ist also dekorreliert (Abb. 14 B). Die aus dem festen Fenster berechnete Autokorrelationsfunktion des Prädiktionsfehlers

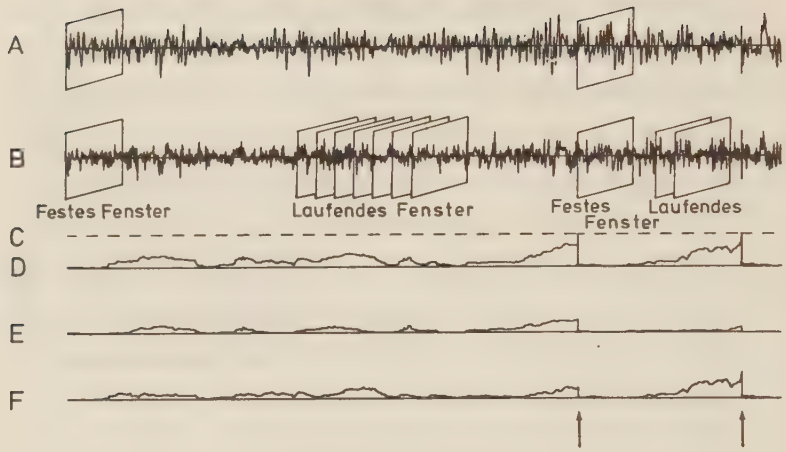

Abb. 14: Erkennung von Segmentgrenzen
A: EEG-Signal; aus dem Fenster wird das Prädiktor-Polynom bestimmt.
B: Prädiktionsfehler; aus dem festen und einem laufenden Fenster werden Autokorrelationsfunktionen berechnet und miteinander verglichen.
C: Schwelle zur Grenzenerkennung.
D: Spektrales Fehlermaß $\hat{E}$ (Gleichung 16).
E: Leistungsanteil von $\hat{E}$.
F: Frequenzanteil von $\hat{E}$.
Die Pfeile markieren erkannte Segmentgrenzen.

ist also bis auf den ersten Wert identisch Null (vgl. Abb. 11, rechts unten). Solange die aus einem laufenden Fenster berechnete Autokorrelationsfunktion des Prädiktionsfehlers von der des festen Fensters nicht abweicht, haben sich die statistischen Parameter des Signals nicht geändert. Wenn sie erheblich voneinander abweichen, wird

an diesen Stellen eine Segmentgrenze erkannt. Als Kriterium bietet es sich an, die Autokorrelationskoeffizienten zu quadrieren und aufzusummieren.

Prinzipiell sind zwei verschiedene Fälle bei der Grenzenerkennung möglich (Leistungsunterschiede sollen außer Betracht bleiben): Das Prädiktor-Polynom wird aus einem Signalteil mit langsamen Frequenzen berechnet; Grenzen müssen erkannt werden, sobald im Signal schnellere Frequenzen auftreten. Dieser Fall ist in Abb. 15 gezeigt. Der umgekehrte Fall, bei dem einem Segment mit schnellen Frequenzen eines mit langsamen folgt, ist in Abb. 16 dargestellt.

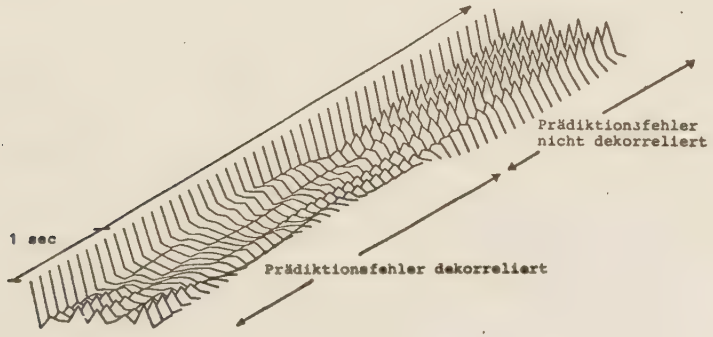

Abb. 15: Autokorrelationsfunktionen des Prädiktionsfehlers, die für jeden 5. Wert neu bestimmt wurden. Wenn der Prädiktionsfehler dekorreliert ist, bleibt $\hat{E}$ klein. Erkennung eines Segmentes mit einer schnelleren Frequenz.

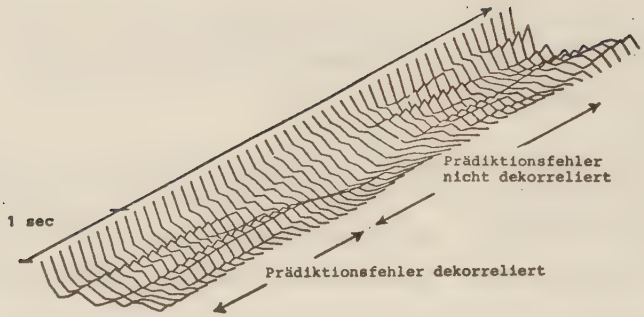

Abb. 16: Autokorrelationsfunktionen des Prädiktionsfehlers. Erkennung eines Segmentes mit einer langsameren Frequenz.

Das Kriterium zur Erkennung von Segmentgrenzen kann als Differenz zwischen zwei Spektren interpretiert werden.

$$E = \frac{1}{2\pi} \int_{-\pi}^{\pi} (P_L(\omega) - P_F(\omega))^2 \, d\omega \qquad (14)$$

wobei $P_F(\omega)$ das Spektrum des durch das feste Fenster (Index F) gesehenen Signals ist, und $P_L(\omega)$ das Spektrum des durch ein laufendes Fenster (Index L) gesehenen. Wenn die quadrierte Differenz zwischen beiden eine bestimmte Schwelle überschreitet, wird eine Segmentgrenze erkannt. E kann als Fehlermaß interpretiert werden, das die Differenz zweier Spektren mißt.

Falls wir nach Gleichung (14) vorgingen, würden die Absolutwerte der Spektren verglichen, was i.a. eine stärkere

Wichtung der langsamen gegenüber den schnellen Frequenzen mit sich brächte. Deshalb muß das Signal vor Berechnung von E möglichst zu einem weißen Rauschen transformiert werden ("prewhitening"), wozu wir das inverse Filter benutzen. Das spektrale Fehlermaß E wird daher nicht aus dem Originalsignal sondern aus dem Prädiktionsfehler bestimmt; $P_L(\omega)$ und $P_F(\omega)$ sind die quadratischen Spektren des Fehlersignals (Abb. 14 B). Die Spektren können durch die Autokorrelationsfunktion dargestellt werden:

$$P(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(\tau) \cos(\omega\tau)\, d\tau$$

oder im diskreten Fall:

$$P_k = \frac{1}{2\pi} \left(R_o + 2 \sum_{k=1}^{\infty} R_k \cos(\omega k)\right)$$

Damit erhalten wir für E:

$$E = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \left(R_{o,L} + 2 \sum_{k=1}^{\infty} R_{k,L} \cos(\omega k) - R_{o,F} \underbrace{-2 \sum_{k=1}^{\infty} R_{k,F} \cos(\omega k)}_{\equiv 0}\right)^2 d\omega$$

$R_{i,L}$ und $R_{i,F}$ sind die Autokorrelationsfunktionen des Prädiktionsfehlers im laufenden, bzw. im festen Fenster. Der letzte Term ist identisch Null, da der Prädiktionsfehler im festen Fenster weißes Rauschen ist. Wir erhalten:

$$E = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \left(R_{o,L} - R_{o,F} + 2 \sum_{k=1}^{\infty} R_{k,L} \cos(\omega k)\right)^2 d\omega$$

Nach Ausmultiplikation und Integration erhalten wir als endgültiges Ergebnis:

$$E = \frac{1}{2\pi}\left[(R_{o,L} - R_{o,F})^2 + 2\sum_{k=1}^{\infty} R_{k,L}^2\right] \qquad (15)$$

Zur Unterscheidung nennen wir den ersten Term Leistungsanteil, den zweiten Term Frequenzanteil. Um die Berechnung von E möglichst schnell durchführen zu können, darf die Summe in Gleichung (15) nur wenige Glieder umfassen. Nach unseren Erfahrungen reicht es aus, nur drei Glieder zu berücksichtigen. Das Fehlermaß $\hat{E}$ liefert dann nur näherungsweise den Unterschied zweier Spektren:

$$\hat{E} = \left(\frac{R_{o,L}-R_{o,F}}{R_{o,L}}\right)^2 + 2\sum_{k=1}^{3} r_{k,L}^2 \text{ mit } r_{k,L} = \frac{R_{k,L}}{R_{o,L}} \qquad (16)$$

Da $\hat{E}$ für jeden Wert der Zeitreihe berechnet wird, muß die Autokorrelationsfunktion des durch das laufende Fenster gesehenen Prädiktionsfehlers für jeden Wert neu bestimmt werden. Die beiden Anteile nach Gleichung (16) sind in Abb. 14 E und F aufgetragen. Eine Segmentgrenze (Pfeil) wird erkannt, wenn die Summe aus beiden Anteilen (Abb. 14 D) eine Schwelle (Abb. 14 C) überschreitet. Die erste Segmentgrenze in Abb. 14 wird gesetzt, da das Segment eine rechnerbedingte Maximallänge erreicht hat, die zweite Grenze wird im wesentlichen wegen der Änderung des Frequenzanteiles erkannt (Abb. 14 F). Im allgemeinen ändern sich beide Anteile gleichzeitig.

Wie sollen die verschiedenen bei der Segmentgrenzenbestimmung auftretenden Parameter gesetzt werden? Die Länge des Signals zur Berechnung der zur Lösung der Normalgleichungen benötigten Autokorrelationsfunktion muß mindestens so groß sein, wie die größte im Signal auftretende Periode. Frequenzen kleiner .8 Hz werden hergefiltert, da sie für diagnostische Zwecke uninteressant sind; in diesem Bereich treten häufig Artefakte auf. Die Mindestlänge des EEGs muß 1.25 sec betragen. Eine Länge von 2 sec (oder 100 Punkten bei 20 msec Abtastrate) hat sich als geeignet herausgestellt (diese Länge entspricht etwa der Fensterlänge in Abb. 14). Als Länge des Prädiktor-Polynoms ist, wie oben gezeigt wurde, p=8 ausreichend.

Die Autokorrelationsfunktion des Prädiktionsfehlers, die zur Berechnung von $\hat{E}$ gebraucht wird, wird ebenfalls aus 100 Werten bestimmt. Sie wird nur einmal am Anfang eines Segmentes für das feste Fenster ausgerechnet. Die Koeffizienten des um einen Wert verschobenen laufenden Fensters können wir berechnen, indem wir bei jedem Koeffizienten jeweils einen Wert abziehen und einen neuen hinzuaddieren:

$$R_o^{n+1} = R_o^n - s_n^2 + s_{n+N+1}$$

$$R_1^{n+1} = R_1^n - s_n\, s_{n+1} + s_{n+N}\, s_{n+N+1} \quad \text{etc.}$$

N: Länge des Prädiktionsfehlers

Am schwierigsten ist es, die Lage der Schwelle festzulegen, die $\hat{E}$ überschreiten muß, damit eine Segmentgrenze

erkannt wird. Da $\hat{E}$ jedoch aus dem Prädiktionsfehler bestimmt wird, der im festen Fenster ein weißes Rauschen darstellt, ist zu erwarten, daß eine einmal gefundene Schwelle auch völlig andere Signale gut segmentiert. Bei unseren Versuchen hat sich herausgestellt, daß Segmentgrenzen richtig erkannt werden, wenn $\hat{E}$ einen Wert überschreitet, der zwischen .5 und .75 liegt. Zur Begründung dieses Wertes werden die beiden Anteile von $\hat{E}$ getrennt betrachtet:

a) Leistungsanteil

Wenn sich die mittlere Amplitude vom laufenden zum festen Fenster um den Faktor 2 ändert, liefert dieser Anteil den Wert 9, der erheblich über dem Schwellwert von .5 liegt. Ändert sich umgekehrt die Amplitude vom festen zum laufenden Fenster um den Faktor 2, erhalten wir einen Wert von .56, der in der Größe des angegebenen Schwellwertes liegt. Um diesen Unterschied auszugleichen, könnte der Leistungsanteil, statt durch $R_{o,L}$ dividiert zu werden, durch das Maximum von $R_{o,L}$ und $R_{o,F}$ geteilt werden. Wir erhalten damit zwar gute Segmentierungsergebnisse, jedoch kann $\hat{E}$ nicht mehr einfach als spektraler Fehler interpretiert werden.

b) Frequenzanteil

Diesen Anteil abzuschätzen, ist erheblich schwieriger. Wir müssen aber zumindest fordern, daß der Schwellwert nicht durch zufällige Schwankungen überschritten werden darf. Wenn sich der Prädiktionsfehler als normalverteiltes,

weißes Rauschen darstellen läßt, beträgt der mittlere Fehler eines einzelnen Autokorrelationskoeffizienten $\sigma/\sqrt{N}$ (Jenkins, Watts 1969). Da der Prädiktionsfehler den Mittelwert Null hat, kann E_p (Gleichung 9) auch als dessen Varianz interpretiert werden. E_p ist für p=8 i.a. kleiner als 0.1 (vgl. Abb. 13). Der mittlere Fehler liegt also in der Größenordnung von 0.3/10 = 0.03 für N=100. Der mittlere Fehler der quadrierten Autokorrelationskoeffizienten liegt bei 0.06 und ist damit noch eine Größenordnung kleiner als die angegebene Schwelle. Daß diese durch zufällige Schwankungen überschritten wird, ist daher ziemlich ausgeschlossen.

Einige Segmentationsbeispiele sind in Abb. 17 zu sehen.

2.2.2 Erkennung von Transienten

Kurzzeitige Veränderungen im EEG, wie Spitzen oder scharfe Wellen, führen zu großen Werten des Prädiktionsfehlers, da sie aus den schon bekannten Werten des Signals nicht vorhergesagt werden können (Abb. 18, Mitte). Wir gehen davon aus, daß sich das EEG durch eine Grundaktivität mit additiv überlagerten Transienten beschreiben läßt. Zur Bestimmung von Segmentgrenzen messen wir die Veränderungen in der Grundaktivität, indem wir aus dem Prädiktionsfehler Autokorrelationsfunktionen bestimmen. Wenn wir diese aus einem wie in Abb. 18, Mitte, gezeigten Prädiktionsfehler bestimmen wollten, erhielten wir nur

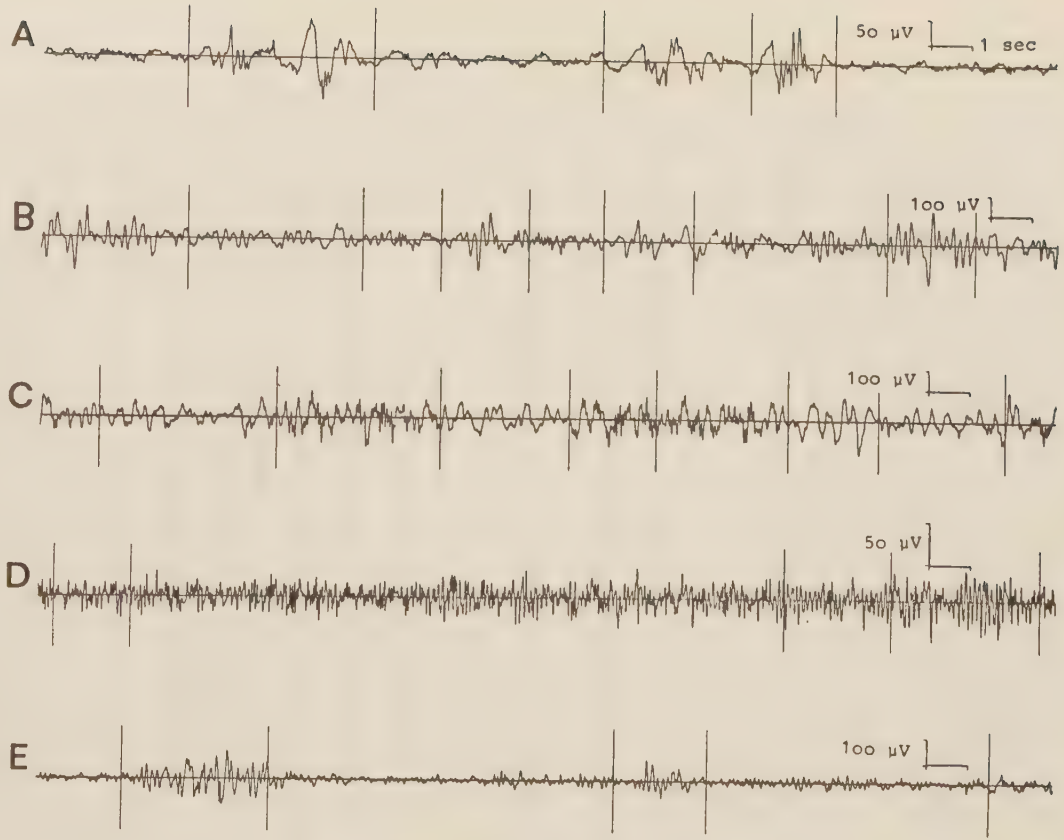

Abb. 17: Automatische Segmentierung von EEGs. A: Kleinstkind, Schlaf; B: Kind, 7 Jahre, Schlafstadium 1; C: Kind, 8 Jahre, Schlafstadium 3; D: normales EEG eines Erwachsenen; E: Erwachsener, EEG mit Paroxysmen.

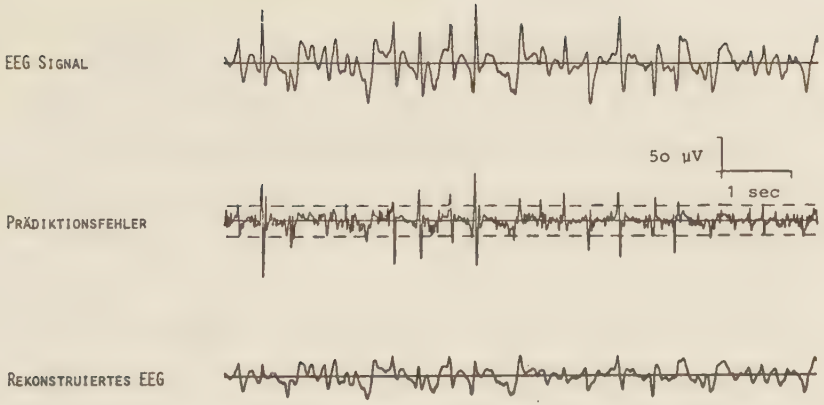

Abb. 18: Erkennung von Transienten. Die über einer Schwelle (Mitte, gestrichelte Linien) liegenden Werte des Prädiktionsfehlers werden abgeschnitten. Nur der Teil innerhalb der Schwelle wird zur Berechnung von $\hat{E}$ benutzt. Die Grundaktivität ohne Transiente kann rekonstruiert werden (unten), sie entspricht dem zur Segmentierung benutzten Signal.

eine sehr schlechte Schätzung des Spektrums der Grundaktivität und auch nur schlechte Segmentgrenzen. Um dieses zu vermeiden, wird der Prädiktionsfehler vor Bestimmung der Autokorrelationsfunktion geclippt, d.h. die Spitzen werden abgeschnitten und nur das übriggebliebene Signal wird zur Grenzenbestimmung benutzt (Abb. 18, Mitte). Wenn wir annehmen, daß sich der Prädiktionsfehler als normalverteiltes weißes Rauschen darstellen läßt, und fordern, daß die Wahrscheinlichkeit zum Überschreiten der Schwelle höchstens p=0.001 betragen soll, muß die Schwelle etwa den Wert 3 haben, der noch mit der jeweiligen Varianz des Prädiktionsfehlers skaliert werden muß. Zur

Veranschaulichung kann aus dem geclippten Fehlersignal die Grundaktivität rekonstruiert werden (Abb. 18, unten) und mit dem Originalsignal verglichen werden. Die Transienten werden bei diesem Verfahren weitgehend eliminiert.

Da die Transienten diagnostisch bedeutungsvoll sein können, müssen sie zur Weiterverarbeitung abgespeichert werden; jedoch wird nicht jede kurzzeitige Veränderung berücksichtigt, sondern nur, wenn die Summe aus drei benachbarten quadrierten Werten des Prädiktionsfehlers einen Schwellwert überschreitet. Diese Schwelle kann man abschätzen (Lopes da Silva et al. 1975): Wenn der Prädiktionsfehler normalverteiltes weißes Rauschen ist, können Vertrauensgrenzen berechnet werden, und zwar gehorcht der Testwert

$$\frac{\varepsilon_{n-1}^2 + \varepsilon_n^2 + \varepsilon_{n+1}^2}{E_p}$$

einer χ^2-Verteilung mit 3 Freiheitsgraden. Vertrauensgrenzen, die z.B. von 95% aller Testwerte nicht überschritten werden, lassen sich damit angeben. Nach unseren Erfahrungen scheinen diese Vertrauensgrenzen zu niedrig zu liegen, da sehr viele Transiente erkannt werden. Die endgültige Schwelle zur Erkennung von Transienten hängt in starkem Maße von der Fragestellung ab und muß den klinischen Forderungen angepaßt werden.

2.2.3 Segmentation und Transientenerkennung für mehrere Kanäle

Häufig müssen mehrere Zeitreihen gleichzeitig untersucht werden. So werden bei EEG-Aufnahmen bis zu 16 Ableitungen gleichzeitig gemacht, mit denen Seitenvergleiche zwischen linker und rechter Hirnhälfte durchgeführt werden können. Um diese Vergleiche auszuführen, müssen Segmentgrenzen in allen Ableitungen gleichzeitig erkannt werden. Es bietet sich an, daß Fehlermaß $\hat{E}$ zu einem Zeitpunkt nT für jede Ableitung einzeln zu berechnen, aufzusummieren und mit einer Schwelle zu vergleichen, sodaß starke Änderungen in einer Ableitung oder leichte in mehreren zu Segmentgrenzen führen können (Abb. 19).

Segmentation und Transientenerkennung führen auch bei völlig anderen stochastischen Signalen, wie geomagnetischen Pulsationen, zu brauchbaren Ergebnissen (Abb. 20). Der einzige Unterschied zur Analyse von EEGs besteht darin, daß die Fensterlängen nach Abb. 14 der größten vorhandenen Periode angepaßt werden müssen. Änderungen im Signal erfolgen häufig nicht so abrupt wie bei EEGs, sondern sind kontinuierlicher, was zu einer größeren Zahl von Segmenten führt.

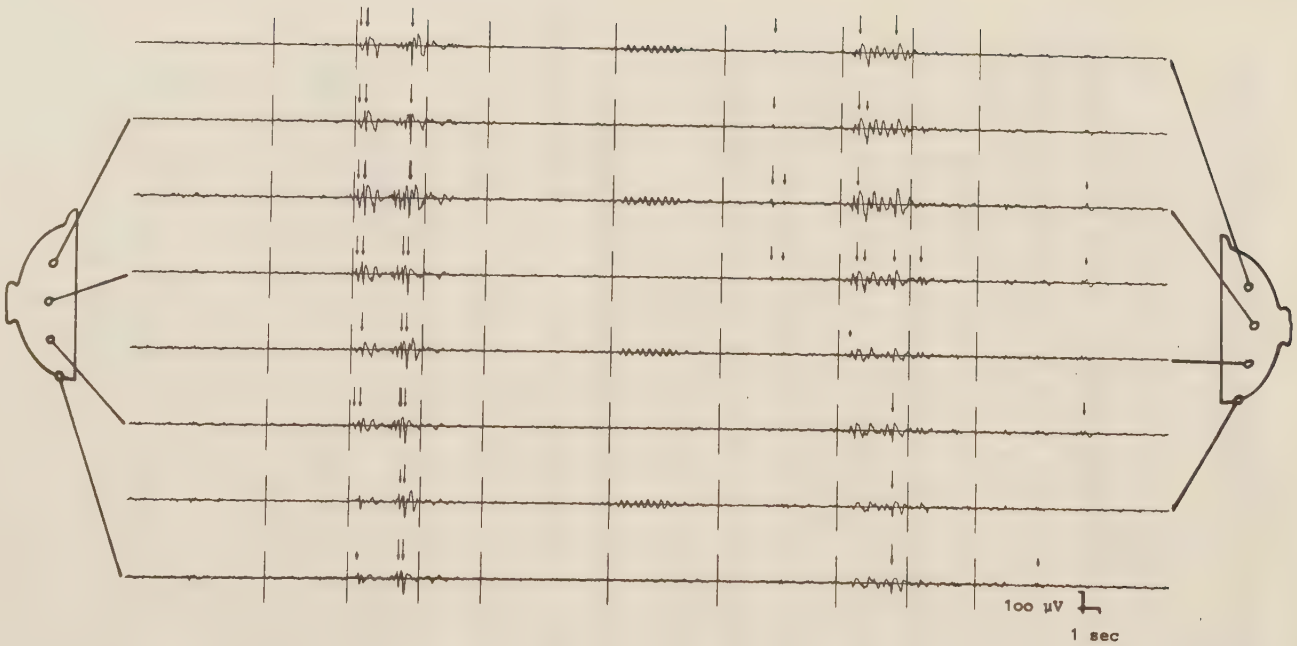

Abb. 19: Gleichzeitige Erkennung von Segmenten und Transienten in 8 Ableitungen. Unterschiede zwischen beiden Schädelhälften sind leicht feststellbar (vgl. Segment 6). Erkannte Transiente sind durch Pfeile gekennzeichnet. Wenn Transiente erkannt sind, werden die entsprechenden Werte des Prädiktionsfehlers abgespeichert (vgl. Abb. 18). Die aus diesen Werten berechnete Länge der Pfeile ist ein Maß für die Größe und Steilheit eines Transienten.

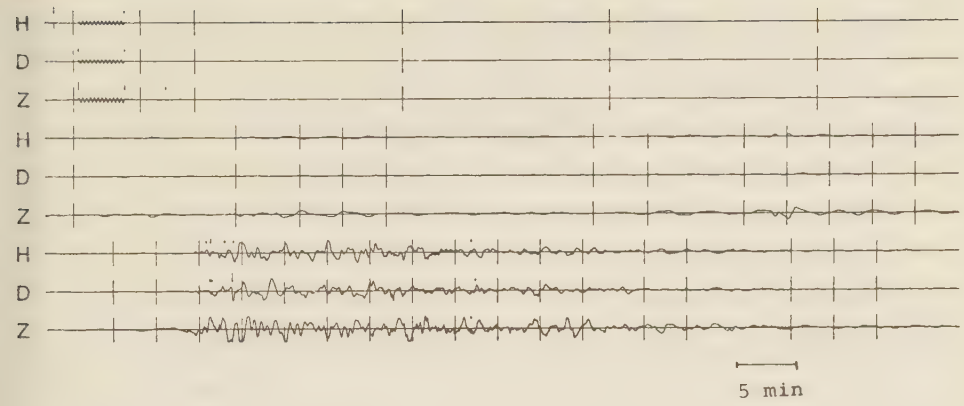

Abb. 20: Segmentation und Erkennung von Transienten bei geomagnetischen Pulsationen. Jede der 3 Registrierungen dauert etwas länger als eine Stunde. Man beachte die Übersteuerung in der H-Komponente der ersten Registrierung und deren Erkennung als Transient. Station: Nordkap, 29.6.1975
Erste Registrierung: 10.00 Uhr; zweite Registrierung: 15.00 Uhr; dritte Registrierung: 20.00 Uhr.

2.2.4 Güte der Merkmale

Nach Erkennung von Transienten werden die entsprechenden Werte des Prädiktionsfehlers abgespeichert, nach Erkennung einer Grenze wird aus dem gesamten Segment eine Autokorrelationsfunktion berechnet und zusammen mit der Segmentlänge abgespeichert. Um ganz sicher zu gehen, daß die Spektren der Segmente gut repräsentiert werden, werden 11 Autokorrelationskoeffizienten bestimmt.

Die Güte der Merkmale kann qualitativ überprüft werden:

Wir wissen, daß der Prädiktionsfehler, falls die Länge des Prädiktor-Polynoms richtig gewählt ist, ein flaches Spektrum hat. Wenn wir das aus der abgespeicherten Autokorrelationsfunktion berechnete, invertierte Prädiktor-Polynom auf künstlich generiertes weißes Rauschen anwenden, müssen wir ein Signal erhalten, das dem ursprünglichen Signal ähnlich ist. Aus der Ähnlichkeit können wir qualitativ abschätzen, ob die abgespeicherte Autokorrelationsfunktion das Spektrum des Originalsignals gut repräsentiert. Abb. 21 zeigt einige Original-EEGs (jeweils oben) sowie die aus weißem Rauschen simulierten Signale (unten). Diese Beispiele wurden Ärzten vorgelegt, die nicht in der Lage waren, die echten von den simulierten EEGs zu unterscheiden, und die der Meinung waren, daß die diagnostische Information in beiden identisch sei.

Da jedes Segment unabhängig von seiner Länge durch einen gleich langen Merkmalssatz beschrieben wird, hängt die Datenreduktion von der Information des Signals selbst ab. Die Datenreduktion liegt, falls keine Transienten auftreten, etwa zwischen 30:1 und 100:1.

Wenn Transienten vorhanden sind, werden die über einer gewissen Schwelle liegenden Werte des Prädiktionsfehlers (s.o.) sowie deren zeitliche Lage gespeichert. Mit dem invertierten Prädiktor-Polynom kann daher die zeitliche Struktur sowie die Lage der Transienten rekonstruiert werden (Abb. 22, Mitte). Darüberhinaus kann

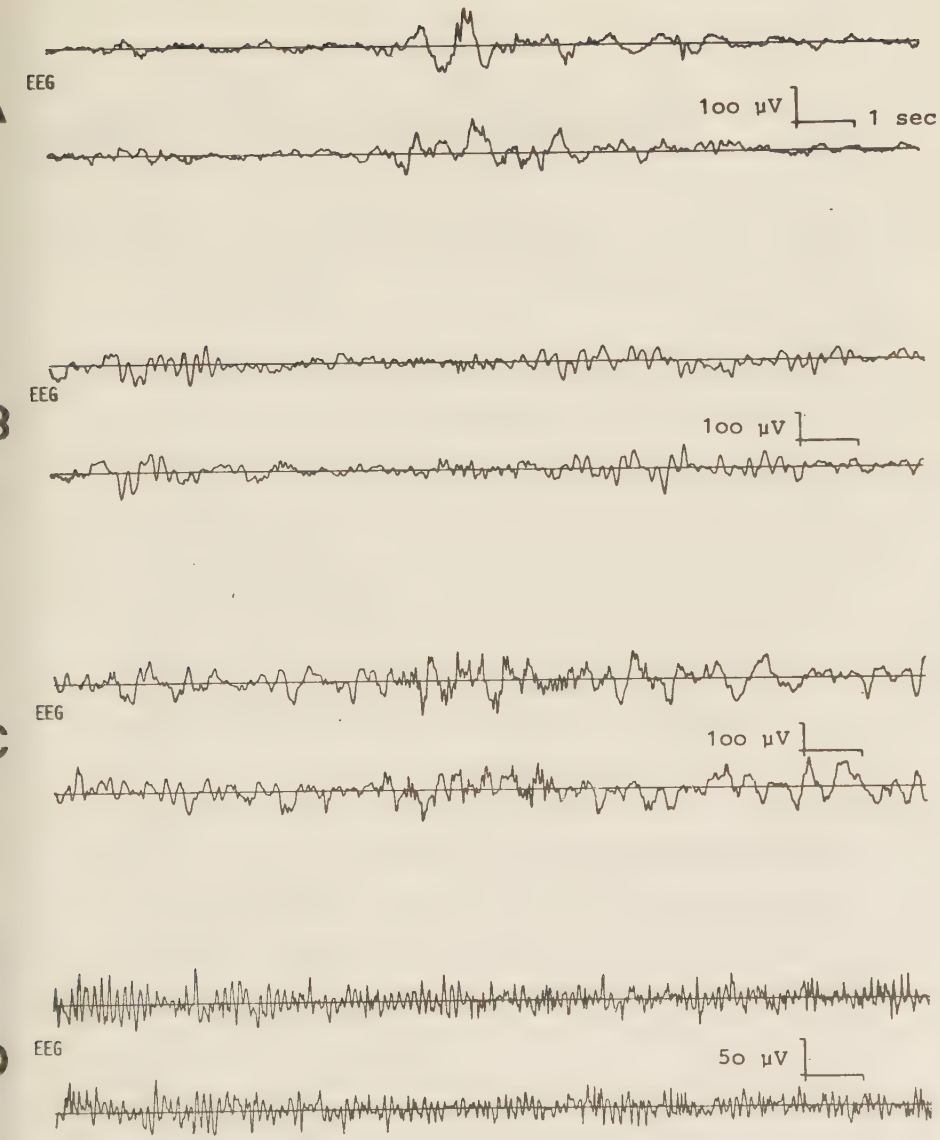

Abb. 21: Simulation von EEGs aus den abgespeicherten Merkmalen. Das EEG (jeweils oben) und das simulierte Signal (unten) können als zwei Realisierungen desselben stochastischen Prozesses aufgefaßt werden. A: Kleinstkind, Schlaf; B: Kind, 7 Jahre, Schlafstadium 1; C: Kind 8 Jahre, Schlafstadium 3, D: normales Erwachsenen-EEG.

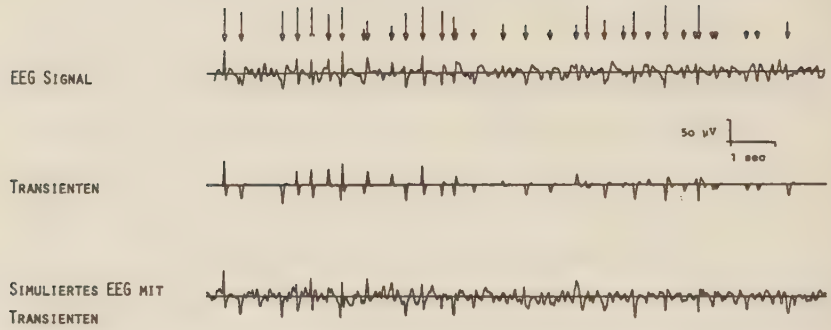

Abb. 22: Simulation eines EEGs mit Transienten (oben); die Pfeile sind ein Maß für die Größe und Steilheit der erkannten Transienten. Aus den abgespeicherten Werten des Prädiktionsfehlers kann die Form der Transienten rekonstruiert werden (Mitte). Durch Addition der simulierten Grundaktivität erhält man ein dem Original-EEG vergleichbares Signal (unten).

durch Addition der simulierten Grundaktivität der reale Bezug von Transienten wiedergewonnen werden (Abb. 22, unten). Die Datenreduktion hängt stark von der Häufigkeit der Transienten ab; sie liegt bei dem gezeigten Beispiel bei etwa 10:1.

3. Verfahren zur Klassifizierung

Die im letzten Kapitel beschriebenen Verfahren zur Merkmalsgewinnung bilden erst einen ersten, wenn auch sehr wesentlichen Teil eines automatischen Systems zur EEG-Analyse. Für den Arzt sind die aus Segmenten und Transienten gewonnenen Merkmale noch nicht interpretierbar. Die meisten automatischen Systeme zur EEG-Analyse bleiben jedoch auf dieser Stufe stehen, z.B. werden als endgültige Ausgabe Verläufe von Spektren über Stunden angegeben, ohne daß der Arzt entscheiden kann, wann sich die Spektren signifikant ändern. Die relativ klar verständlichen EEGs werden in Merkmale umgewandelt, die dem Arzt häufig gänzlich unverständlich sein müssen. Ein automatisches System zur EEG-Analyse darf nicht auf der Stufe der Merkmalsgewinnung stehenbleiben; vielmehr müssen die sich anschließenden Klassifikationsverfahren die mathematischen Merkmale in eine dem Arzt verständliche Ausgabe umwandeln.

Wenn nach der Merkmalsgewinnung Segmente und Transiente erkannt sind (vgl. Abb. 19), müssen im nächsten Schritt ähnliche Merkmalssätze zusammengefaßt werden. Da der Frequenzgehalt der Segmente a-priori nicht bekannt ist, bieten sich Klassifikationsverfahren an, die (fast) keine a-priori Information benötigen, die allein aus den vorhandenen Messungen automatisch Klassen bilden können. Die Segmente einer Klasse können anschließend durch ein repräsentatives Segment dargestellt werden. Verfahren zur automatischen Klassifikation von Segmenten (Clusteranalyse) werden in

Kapitel 3.1 behandelt.

Der Stufe der automatischen Klassifikation muß die Klassifikation des gesamten EEGs nach klinischen Gesichtspunkten, wie die Zuordnung zu Schlafstadien, Altersklassen oder Krankheiten folgen. Da hier die verschiedenen Klassen a-priori bekannt sind, können diskriminanzanalytische Verfahren benutzt werden. Auf Grund von vorgegebenen Klassen werden Entscheidungsflächen (Diskriminanzfunktionen) berechnet, mit deren Hilfe unbekannte Merkmalssätze klassifiziert werden können. Verfahren der Diskriminanzanalyse werden in Kapitel 3.2 behandelt.

Auch Transienten müssen im Prinzip wie Segmente klassifiziert und zusammengefaßt werden. Da sich a-priori etwa fünf klinisch relevante Klassen definieren lassen (Cavazza et al. 1974), werden nicht die Verfahren zur automatischen Klassifikation benötigt. Darüberhinaus ist die Klassifikation von Transienten mehr ein klinisches Problem. Wir wollen im Folgenden nicht mehr darauf eingehen.

Die Auswahl der Merkmale wird weniger von ihrer physikalischen Deutbarkeit, als von den mit ihnen erzielbaren Klassifikationsergebnissen bestimmt. Es werden verschiedene Merkmalssätze auf ihre Klassifikationsgüte hin untersucht. Einige quantitative und qualitative Verfahren sind in Kapitel 3.3 dargestellt.

Zur Illustration der Klassifikationsverfahren werden wir

EEGs benutzen, bei denen allein schon das Auftreten und die Häufigkeit bestimmter Segmente Hinweise für eine Diagnose geben, sowie ein Material von 340 EEGs 4-, 6-, 10- und 12-jähriger Kinder.

3.1 Automatische Klassifikation von Segmenten

Zur automatischen Klassifikation (Clusteranalyse) ist eine große Zahl verschiedener Verfahren entwickelt worden. Das Grundproblem besteht darin, in einer Menge von Objekten homogene Klassen zu entdecken, wobei "ähnliche Objekte möglichst zu gleichen, unähnliche Objekte aber zu verschiedenen Klassen gehören sollen" (Bock 1974). Alle Verfahren müssen die Ähnlichkeit zwischen verschiedenen Objekten messen und/oder die Homogenität der einzelnen Klassen objektiv beschreiben. Als Ähnlichkeitsmaße zwischen zwei Objekten werden Distanzen benutzt, wie der einfache euklidsche Abstand oder Verallgemeinerungen davon, wie die Mahalanobis-Distanz, die eine eventuelle Abhängigkeit der Merkmale mit berücksichtigt.

Die Homogenität einer Klasse kann gemessen werden, wenn eine mittlere Distanz aus allen Merkmalen gebildet wird oder wenn die Summen der quadratischen Abweichungen von einem Klassenschwerpunkt (Varianz) berechnet wird. Es existieren noch eine große Zahl anderer Ähnlichkeits- und Homogenitätsmaße, die allerdings häufig nur bei speziellen Fragestellungen oder besonderen Merkmalen sinnvoll angewandt werden können.

Die Verfahren zur automatischen Klassifikation lassen sich in drei Gruppen unterteilen: Solche, die von einer fest vorgegebenen Klassenzahl ausgehen, und welche, die die Homogenität der Klassen a-priori definieren. Bei der automatischen Klassifikation von Segmenten sind beide Arten der Information a-priori nicht bekannt; als einzige Information steht zur Verfügung, daß die Zahl der Segmentklassen eines EEGs nicht zu groß sein soll, z.B. nicht größer als sechs. Beide Klassifikationsstrategien sind also ungeeignet. Man kann diese Schwierigkeiten umgehen, indem man schrittweise neue Klassen erzeugt, wobei die Homogenität und die Zahl der Klassen von Schritt zu Schritt abnehmen. Diese Klassifikationsart nennt man hierarchische Gruppierung. Wir benutzen ein Verfahren, das bei m Segmenten von m Klassen ausgeht und bei jedem Schritt aus den beiden ähnlichsten Klassen eine neue erzeugt, bis schließlich nur noch eine Klasse, die alle Segmente umfaßt, übrig bleibt. Die Zusammenfassung von ähnlichen Klassen kann graphisch als Dendrogramm dargestellt werden.

3.1.1 Die Methode von Ward

Bei Beginn der hierarchischen Gruppierung sind m Klassen vorhanden, von denen jede ein einziges Segment mit p Merkmalen enthält. Bei jedem der m-1 Fusionsschritte werden die beiden ähnlichsten Klassen zusammengefaßt. Bei m Klassen müssen daher $\binom{m}{2}$ mögliche Fusionen verglichen werden. Ward (1963) benutzt als Homogenitätsmaß die Summen

der quadrierten Distanzen jedes Merkmales einer Klasse zum Schwerpunkt der zugehörigen Klasse ("error sum of squares", ESS). Bei jedem Schritt werden die beiden Klassen fusioniert, die den kleinsten Zuwachs von ESS liefern.

Angenommen, die beiden Klassen P und Q sollen vereinigt werden. Der Vektor $\underline{x}_i$ verbinde den i-ten Merkmalssatz mit dem Schwerpunkt der Klasse P und n_P sei die Zahl der Merkmalssätze in Klasse P; $\underline{x}_j$ und n_Q seien die entsprechenden Werte für Klasse Q. Wards Fehlermaß ESS ist dann für Klasse P:

$$ESS_P = \sum_{i=1}^{n_P} \underline{x}_i^2$$

und für Klasse Q:

$$ESS_Q = \sum_{j=1}^{n_Q} \underline{x}_j^2 \ .$$

Der Vektor $\underline{d}$ verbinde die beiden Schwerpunkte von P und Q. ESS ist also für die aus P und Q gebildete Klasse A:

$$ESS_A = \sum_{i=1}^{n_P} \left(\underline{x}_i - \frac{n_Q}{n_P+n_Q}\,\underline{d}\right)^2 + \sum_{j=1}^{n_Q} \left(\underline{x}_j - \frac{n_P}{n_P+n_Q}\,\underline{d}\right)^2$$

Der Zuwachs der totalen Fehlersumme bei Fusion von P und Q ist:

$$\Delta E_{PQ} = ESS_A - ESS_P - ESS_Q$$

Durch Einsetzen der entsprechenden Fehlersummen erhält man (Wishart 1969):

$$\Delta E_{PQ} = \frac{n_P \, n_Q}{n_P + n_Q} \, d^2 \qquad (17)$$

$n_{P,Q}$: Größe der zu fusionierenden Klassen P und Q

d: Entfernung zwischen ihren Schwerpunkten.

Die beiden Klassen, für die E_{PQ} minimal ist, werden agglomeriert. Aus Gleichung (17) ist ersichtlich, daß bevorzugt Klassen mit wenig Merkmalssätzen fusioniert werden.

3.1.2 Klassifikationsbeispiele

Als Beispiel zur automatischen Klassifikation benutzen wir ein EEG, bei dem ein Arzt nur zwei deutlich verschiedene Segmentklassen unterscheiden würde, und zwar treten nur Segmente auf mit einer vorherrschenden Frequenz von 7 - 8 Hz oder solche mit einer zusätzlichen schnellen 14-Hz-Komponente. Insgesamt sind 20 Segmente mit der langsamen Frequenz und 14 mit der schnelleren Frequenz vorhanden, die Gesamtzahl ist also m = 34. Das zu dem m-1 Fusionsschritten gehörende Dendrogramm ist in Abb. 23 zu sehen. Die 34 Enden des Dendrogramms (unten) repräsentieren die 34 Segmente. Die zeitliche Reihenfolge geht bei dieser Darstellung allerdings verloren; die ersten drei Segmente links sind z.B. das 3., 18. und 21. Segment des EEGs, d.h. diese Segmente sind sehr ähnlich und werden auf einem niedrigen Niveau fusioniert. Die Ordinate ist proportional zur Wardschen Fehlersumme ESS.

Eine dem Signal entsprechende sinnvolle Klassenzahl kann

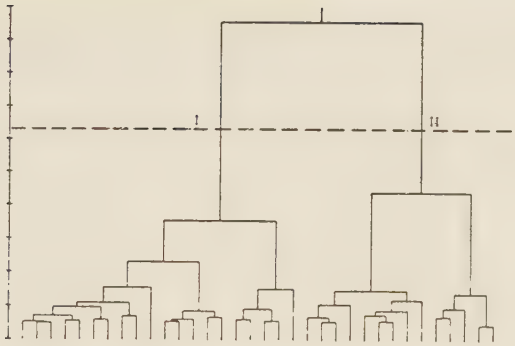

Abb. 23: Automatische Klassifikation von Segmenten. Im Dendrogramm wird die Fusion der jeweils ähnlichsten Segmente dargestellt. Eine Gruppierung, die zu zwei verschiedenen Klassen I und II führt, ist gestrichelt eingezeichnet. Merkmalssatz: Prädiktor-Polynom.

automatisch festgelegt werden, wenn wir die Änderungen von ESS untersuchen, wobei zusätzlich zu fordern ist, daß die Klassenzahl klein sein sollte. Eine vernünftige Gruppierung ist in Abb. 23 gestrichelt eingezeichnet. Sie führt zu zwei Klassen I und II, deren Segmente einzeln in Abb. 24 dargestellt sind.

Ein anderes EEG, das sich ebenfalls gut durch zwei Klassen beschreiben läßt, ist in den Abbildungen 25 und 26 zu sehen. Die im Dendrogramm eingezeichnete Gruppierung in 4 Klassen ist nicht sinnvoll; die Änderung von ESS ist an dieser Stelle zu klein verglichen mit der Änderung bei der letzten Fusion. Die Segmente der Klassen I und II und die der Klassen III und IV (Abb. 26) zeigen daher nur geringe Unterschiede.

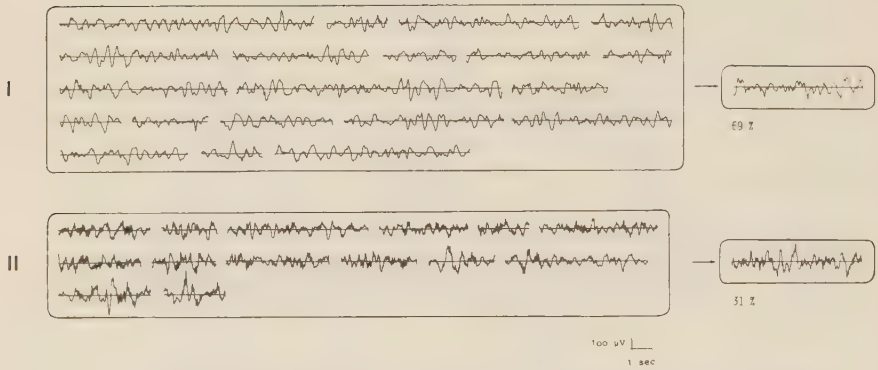

Abb. 24: Darstellung der einzelnen Segmente für die nach Abb. 23 gefundenen Klassen. Repräsentative Segmente und die Gesamtdauer aller Segmente einer Klasse (in %) sind rechts angegeben.

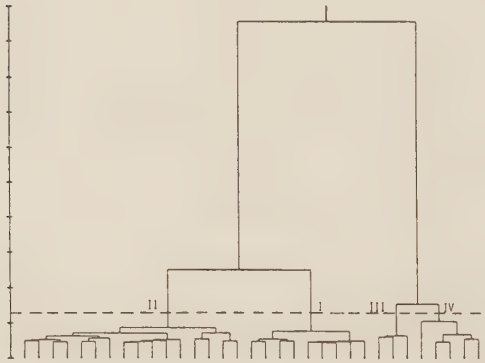

Abb. 25: Automatische Klassifikation von Segmenten mit dem Prädiktor-Polynom. Eine Gruppierung mit vier Klassen ist eingezeichnet.

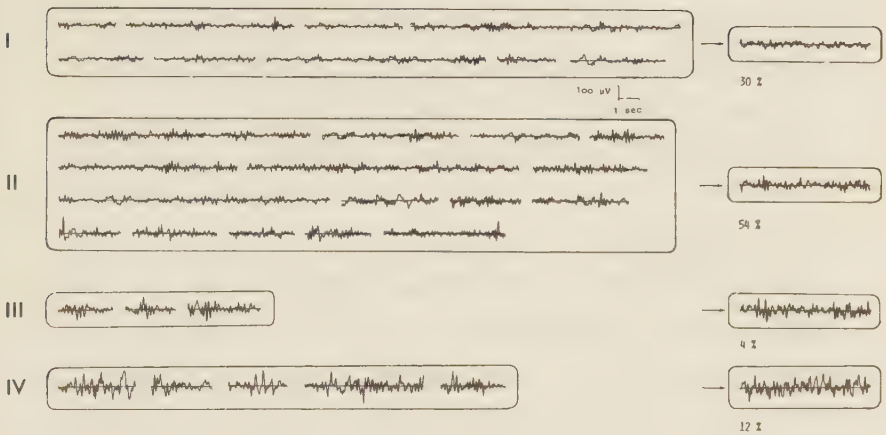

Abb. 26: Darstellung der einzelnen Segmente für die nach Abb. 25 gefundenen Klassen mit den repräsentativen Segmenten.

Nach der automatischen Klassifikation können die Segmente einer Klasse durch ein einziges Segment repräsentiert werden. Durch Mittelung der Autokorrelationsfunktionen der einzelnen Segmente erhalten wir die Autokorrelationsfunktion des repräsentativen Segments, aus der wir - wie in Kapitel 2.2.4 beschrieben - mit Hilfe des invertierten Prädiktor-Polynoms ein repräsentatives Segment simulieren können. Dies ist für jede Klasse in den Abbn. 24 und 26 jeweils rechts dargestellt mit der zeitlichen Dauer (in %) aller Segmente einer Klasse.

Für jede Klasse braucht nur die repräsentative Autokorrelationsfunktion abgespeichert zu werden, sowie eine Zeitreihe, die aus der Länge und der Klassennummer der Segmente

besteht. Beide Informationen können zu einem Rechnerwort zusammengefaßt werden. Lag die Datenreduktion nach der Merkmalsgewinnung zwischen 30:1 und 100:1, so ist sie bei den gezeigten Beispielen um den Faktor 10 größer. Sie liegt in einer Größenordnung von 300:1 bis 1000:1.

Schon diese Klassifikationsstufe kann für eine Diagnose von Bedeutung sein. Es gibt eine Krankheit, bei der Segmente mit schnellen Komponenten (Abb. 24) in großer Anzahl auftreten. Allein aus der Häufigkeit und Länge, die wie in Abb. 24, rechts, dem Arzt gezeigt werden können, ist es möglich, einen Hinweis auf diese Krankheit zu gewinnen.

Ebenso treten sog. Paroxysmen nur bei anomalen EEGs auf (Abb. 26, Klassen III und IV), wobei die Bezeichnung dieser Segmente als Paroxysmen nur dem geschulten Arzt möglich ist. Ein automatisches Analysesystem sollte auch diese Zuordnung durchführen können. Die dazu benötigten Hilfsmittel werden im nächsten Kapitel dargestellt.

3.2 Klassifikation nach vorgegebenen Gesichtspunkten

Nach der automatischen Segmentklassifikation, die unabhängig von klinischen Gesichtspunkten zu erfolgen hat, müssen die gewonnenen Merkmalssätze a-priori bekannten Klassen, wie Altersklassen, Schlafstadien oder Krankheiten, zugeordnet werden.

Als Hypothese nehmen wir an, daß die Lage der Klassen unterscheidbar ist, und zwar daß speziell die Lage der Mittelpunkte verschieden ist (vgl. Abb. 27). Diese Forderung ist notwendig: wenn die klinisch definierten Klassen nicht unterscheidbar wären, könnte keine Diagnose gestellt werden. Das EEG wird der Klasse zugeordnet, zu der es am ähnlichsten ist. Als Zuordnungskriterien können Distanzen oder Wahrscheinlichkeiten benutzt werden. Das EEG wird dann der Klasse zugeordnet, zu der es die kleinste Distanz hat oder zu der es mit der größten Wahrscheinlichkeit gehört.

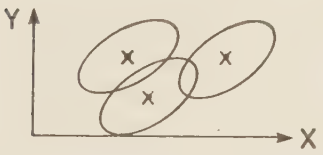

Abb. 27: Darstellung von drei 2-dimensionalen Klassen mit ungleichen Mittelwerten (nach: Cooley, Lohnes 1971).

Zur Lösung der Probleme müssen wir in zwei Schritten vorgehen. Zuerst müssen die Häufigkeitsverteilungen der einzelnen Klassen bestimmt und daraus Entscheidungsflächen - Diskriminanzfunktionen - berechnet werden (Lernphase). Im zweiten Schritt werden die EEGs mit Hilfe der Diskriminanzfunktionen den einzelnen Klassen zugeteilt (Klassifikationsphase). Beide Stufen zusammen werden als Diskriminanzanalyse bezeichnet.

Einzelne Segmente könnten z.B. nach der dominierenden Frequenz klassifiziert werden (Abb. 28). Die Diskriminanz-

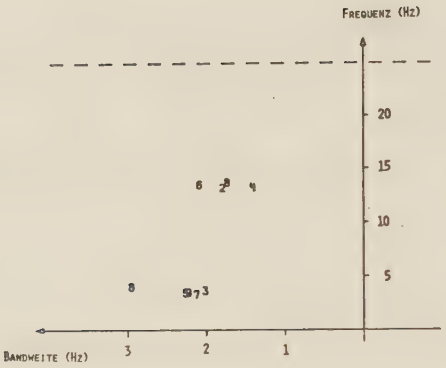

Abb. 28: Klassifikation von Segmenten nach der Frequenz. Jede Zahl repräsentiert ein Segment.

funktion könnte in diesem Fall eine Gerade sein, die parallel zur Abszisse verläuft und die Ordinate bei etwa 8 Hz schneidet.

Sei $\underline{X}_{ki}$ der i-te Merkmalsvektor der k-ten Klasse. Dieser Vektor soll ein EEG-Segment mit p Merkmalen repräsentieren. Wir können die Identität schreiben:

$$\underline{X}_{ki} = \underline{m} + (\underline{m}_k - \underline{m}) + (\underline{X}_{ki} - \underline{m}_k) \; ; \quad \begin{matrix} 1 \leq k \leq g \\ 1 \leq i \leq N_k \end{matrix}$$

g: Zahl der Klassen

N_k: Zahl der Merkmalsvektoren pro Klasse

$\underline{m}$ sei der Schwerpunkt aller Merkmalsvektoren aller Klassen. $\underline{m}_k$ sei der Schwerpunkt der Klasse k. Da i.a. die Abhängigkeit von $\underline{m}$ uninteressant ist, wird $\underline{m}$ von allen Merkmalsvektoren abgezogen:

$$\underline{x}_{ki} = \underline{X}_{ki} - \underline{m} \quad \text{und} \quad \underline{x}_{ki} = (\underline{m}_k - \underline{m}) + (\underline{X}_{ki} - \underline{m}_k)$$

Der erste Term auf der rechten Seite beschreibt verschiedene Klassenlagen, der zweite die Abweichung vom jeweiligen Klassenschwerpunkt. Wenn wir jetzt über alle Quadrate und Kreuzprodukte aller Merkmalsvektoren und Klassen summieren, erhalten wir (Cooley, Lohnes 1971):

$$\sum_{k=1}^{g} \sum_{i=1}^{N_k} \underline{x}_{ki} \underline{x}_{ki}^T = \sum_{k=1}^{g} \sum_{i=1}^{N_k} (\underline{m}_k - \underline{m})(\underline{m}_k - \underline{m})^T + \sum_{k=1}^{g} \sum_{i=1}^{N_k} (\underline{X}_{ki} - \underline{m}_k)(\underline{X}_{ki} - \underline{m}_k)^T \qquad (18)$$

Der Ausdruck auf der linken Seite ist die Summe der Quadrate und Kreuzprodukte aller Abweichungen aller Merkmalsvektoren aller Klassen vom Schwerpunkt $\underline{m}$. Diese Matrix wird in der Literatur i.a. mit $\underline{\underline{T}}$ ("total") abgekürzt. Der erste Ausdruck auf der rechten Seite beschreibt die Abweichung der Klassenschwerpunkte $\underline{m}_k$ vom gemeinsamen Schwerpunkt $\underline{m}$. Er wird mit $\underline{\underline{A}}$ ("among-classes") abgekürzt. Der zweite Ausdruck beschreibt die Abweichungen der Merkmalsvektoren von den einzelnen Klassenschwerpunkten. Er wird mit $\underline{\underline{W}}$ ("within-classes") abgekürzt. Damit können wir Gleichung (18) schreiben:

$$\underline{\underline{T}} = \underline{\underline{A}} + \underline{\underline{W}} \; .$$

3.2.1 Grundlagen der Diskriminanzanalyse

Die der Diskriminanzanalyse zugrunde liegende Idee soll an einem Zwei-Klassen-Problem erklärt werden. Beide Klassen

lassen sich durch p-dimensionale Wahrscheinlichkeitsverteilungen beschreiben. Jedes einzelne Element läßt sich geometrisch als Punkt im p-dimensionalen Raum deuten. Die Elemente beider Klassen sollen orthogonal auf eine Gerade projieziert werden, sodaß die Unterschiede zwischen den Klassen so groß wie möglich, die innerhalb der Klassen so klein wie möglich werden (Abb. 29). Das Problem besteht darin, die Geradengleichung zu bestimmen.

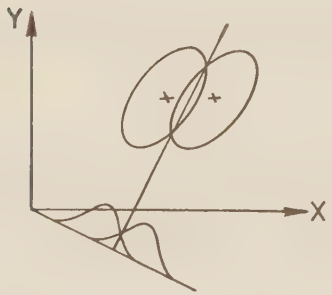

Abb. 29: Geometrische Interpretation der Diskriminanzanalyse (s. Text, nach: Cooley, Lohnes 1971).

Für g Klassen soll entsprechend der p-dimensionale Merkmalsraum orthogonal auf einen (g-1)-dimensionalen Raum projieziert werden, wobei das Verhältnis der Determinante von $\underline{\underline{W}}$ dividiert durch die Determinante von $\underline{\underline{T}}$ minimiert werden soll (Wilks 1962):

$$\Lambda = \frac{|\underline{\underline{W}}|}{|\underline{\underline{T}}|} \qquad (19)$$

Dieses Kriterium führt für zwei Klassen auf folgende Eigenwertgleichung:

$$(\underline{\underline{W}}^{-1}\ \underline{\underline{A}} - \lambda\ \underline{\underline{I}})\ \underline{v} = 0 \qquad (20)$$

λ: Eigenwert
$\underline{v}$: Eigenvektor
$\underline{\underline{I}}$: Einheitsmatrix

Wenn g Klassen vorhanden sind, erhalten wir genau g-1 Eigenwerte und die entsprechenden Eigenvektoren. Allein aus den Eigenwerten kann Λ berechnet werden:

$$\Lambda = \frac{1}{(1+\lambda_1)(1+\lambda_2)\ldots(1+\lambda_{g-1})} = \prod_{k=1}^{g-1} \frac{1}{1+\lambda_k}$$

Wenn wir die Matrix $\underline{\underline{V}}$ der g-1 Eigenvektoren bestimmt haben, können wir die p-dimensionalen Merkmalsvektoren $\underline{x}_{ki}$ transformieren nach:

$$\underline{z}_{ki} = \underline{\underline{V}}\, \underline{x}_{ki} \qquad (21)$$

Die Vektoren $\underline{z}_{ki}$ haben die Dimension g-1. Diese Gleichung wird in der Diskriminanzanalyse als kanonische Transformation bezeichnet. Sie kann interpretiert werden als orthogonale Projektion des p-dimensionalen Merkmalraumes auf einen (g-1)-dimensionalen Raum, wobei Λ in beiden Räumen identisch ist (Wilks 1962). Für die praktische Anwendung ist dies eminent wichtig: Wenn das Eigenwertproblem (20) einmal in der Lernphase gelöst ist, braucht die Klassifikation von unbekannten Merkmalsvektoren nicht mehr im p-dimensionalen sondern nur noch im (g-1)-dimensionalen Raum zu erfolgen.

Mathieu et al. (1976) benutzt dieses Verfahren zur Schlafstadienbestimmung. Er leitet EEGs in vier Kanälen ab und berechnet daraus Vektoren mit je 9 Merkmalen, sodaß p=36 ist. Da sich Schlafstadien in sechs Klassen einteilen lassen,

braucht er nach der kanonischen Transformation die Klassifikation nicht mehr im 36- sondern nur noch im 5-dimensionalen Raum auszuführen.

Die Verfahren der Diskriminanzanalyse sind nur sinnvoll anzuwenden, wenn die Klassen verschiedene Mittelwerte besitzen. Man kann die Nullhypothese, nämlich daß die Mittelwerte gleich sind, mit dem F-Test prüfen, wobei man davon ausgeht, daß sich die Verteilungen der Merkmale durch mehrdimensionale Normalverteilungen beschreiben lassen (Rao 1973). Es gilt als Testwert:

$$F_{n_2}^{n_1} = \left(\frac{1 - \Lambda^{1/s}}{\Lambda^{1/s}}\right)\left(\frac{n_2}{n_1}\right) \qquad (22)$$

mit $n_1 = p(g-1)$

$$n_2 = s\left((N-1) - \frac{p+(g-1)+1}{2}\right) - \frac{p(g-1)-2}{2}$$

$$s = \left(\frac{p^2(g-1)^2-4}{p^2+(g-1)^2-5}\right)^{1/2}$$

p: Zahl der Merkmale
g: Zahl der Klassen
N: Gesamtzahl aller Merkmalsvektoren aller Klassen

3.2.2 Klassifikation

Es sei P_k die a-priori Wahrscheinlichkeit für das Auftreten der k-ten Klasse, $p(\underline{x}|k)$ sei die Wahrscheinlichkeitsdichte von $\underline{x}$ unter der Bedingung, daß $\underline{x}$ zur k-ten Klasse gehöre. Für den Fall, daß alle Merkmalsvektoren sich den bekannten

Klassen zuordnen lassen - es sei also keine Zurückweisung erlaubt - gilt nach der Bayes-Regel (Fukunaga 1972): Der Merkmalsvektor $\underline{x}$ gehöre zur Klasse j, wenn

$$P_j \cdot p(\underline{x}|j) \geq P_i \; p(\underline{x}|i) \text{ für alle } i \neq j \qquad (23)$$

Zur Klassifikation sind Ausdrücke der Form $d_k = P_k p(\underline{x}|k)$ zu bilden und miteinander zu vergleichen. Da es nur auf relative Vergleiche ankommt, können auch monotone Funktionen von d_k, wie $\ln(d_k)$ bestimmt werden (Meyer-Brötz, Schürmann 1970). Um mit Gleichung (23) zu praktischen Ergebnissen zu kommen, wird als Wahrscheinlichkeitsdichte $p(\underline{x}|k)$ häufig eine mehrdimensionale Normalverteilung angenommen. Es sei

$$p(\underline{x}|k) = \frac{1}{(2\pi)^{p/2} |\underline{\underline{K}}_k|^{1/2}} \exp\left(-\frac{1}{2}(\underline{x}-\underline{m}_k)^T \underline{\underline{K}}_k^{-1} (\underline{x}-\underline{m}_k)\right) \qquad (24)$$

$\underline{m}_k$: Schwerpunktsvektor der Klasse k mit p Komponenten

$\underline{\underline{K}}_k$: Kovarianzmatrix der Klasse k.

Um ein anschauliches Bild der Klassen sowie deren Größe zu erhalten, können für jede Klasse Flächen konstanter Wahrscheinlichkeitsdichte berechnet werden. Speziell für eine Wahrscheinlichkeitsdichte nach Gleichung (24) ergeben sich Hyperellipsoide oder im 2-dimensionalen Fall Ellipsen, deren Achsen jedoch i.a. nicht mit der Richtung der Einheitsvektoren übereinstimmen, vielmehr mit diesen einen Winkel ϕ einschließen. Um die Ellipsen graphisch

darstellen zu können, muß eine Hauptachsentransformation der Kovarianzmatrix durchgeführt werden. Die Kovarianzmatrix läßt sich im 2-dimensionalen Fall (p=2) schreiben als:

$$\underline{\underline{K}} = \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} ;$$

da sie symmetrisch ist, gilt:

$$K_{12} = K_{21} .$$

Im 2-dimensionalen Fall führt die Hauptachsentransformation auf eine Eigenwertgleichung mit zwei Eigenwerten λ_1 und λ_2 als Lösung. $\underline{\underline{K}}$ läßt sich dann schreiben:

$$\underline{\underline{K}} = \begin{pmatrix} \lambda_1 & o \\ o & \lambda_2 \end{pmatrix} .$$

Der Winkel ϕ zwischen Einheitsvektor und Hauptachse der Ellipse kann bestimmt werden:

$$tg\phi = \frac{K_{11} - \lambda_1}{K_{12}}$$

Jetzt können wir die Kurven konstanter Wahrscheinlichkeitsdichte z.B. für Klasse 1 angeben (im Folgenden wird statt $\underline{\underline{K}}_1$ nur $\underline{\underline{K}}$ und statt $\underline{m}_1$ nur $\underline{m}$ geschrieben):

$$p(\underline{x}|1) = \frac{1}{2\pi|\underline{\underline{K}}|^{1/2}} \exp\left(-\frac{1}{2}(\underline{x}-\underline{m})^T \underline{\underline{K}}^{-1}(\underline{x}-\underline{m})\right)$$

$$= \frac{1}{2\pi|\underline{\underline{K}}|^{1/2}} \exp\left(-\frac{1}{2}(x_1-m_1, x_2-m_2) \begin{pmatrix} \lambda_1^{-1} & o \\ o & \lambda_2^{-1} \end{pmatrix} \begin{pmatrix} x_1 - m_1 \\ x_2 - m_2 \end{pmatrix}\right)$$

$$= \frac{1}{2\pi|\underline{\underline{K}}|^{1/2}} \exp\left\{-\frac{1}{2}((x_1-m_1)^2\lambda_1^{-1} + (x_2-m_2)^2\lambda_2^{-1})\right\}$$

Logarithmieren und Auflösen ergibt:

$$\frac{(x_1-m_1)^2}{-2\ \ln(2\pi|\underline{\underline{K}}|^{1/2}p(\underline{x}|1))}\lambda_1^{-1} + \frac{(x_2-m_2)^2}{-2\ \ln(2\pi|\underline{\underline{K}}|^{1/2}p(\underline{x}|1))}\lambda_2^{-1} = 1$$

Dies ist die Gleichung einer Ellipse mit den beiden Achsen:

$$\left(\frac{-2\ \ln(2\pi|\underline{\underline{K}}|^{1/2}p(\underline{x}|1)}{\lambda_{1,2}^{-1}}\right)^{1/2}$$

und dem Mittelpunktsvektor $\underline{m}^T = (m_1, m_2)$. Wenn wir für $p(\underline{x}|1)$ einen Wert von 0.01 einsetzen, bedeutet das, daß 99% aller Merkmalsvektoren innerhalb der Ellipse liegen.

Zur Illustration werden die schon früher gefundenen Gruppierungen der Abbn. 24 und 26 untersucht, um zu sehen, inwieweit sich die automatisch gefundenen Klassen auch mit den Methoden der Diskriminanzanalyse trennen lassen. Da im ersten Fall nur zwei Klassen vorhanden sind (Abb. 24) werden die 12-dimensionalen Merkmalsvektoren mit der kanonischen Transformation auf einen 1-dimensionalen Raum, also eine Gerade abgebildet (Abb. 30, Mittelpunkte sind durch Kreuze gekennzeichnet). Beide Klassen sind deutlich

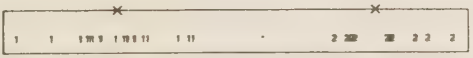

Abb. 30: Darstellung von Segmenten der nach Abb. 23 gefundenen Klassen nach kanonischer Transformation. Beide Klassen sind gut zu trennen. Jede Zahl repräsentiert ein Segment. Die Lage der Mittelpunkte ist durch Kreuze gekennzeichnet.

voneinander zu trennen. Im zweiten Fall (Abb. 26) sind nach der Clusteranalyse vier Klassen vorhanden. Die Merkmalsvektoren werden also in einen 3-dimensionalen Raum projeziert. Die beiden Komponenten, die zu den größten Eigenwerten gehören, sind in Abb. 31 dargestellt.

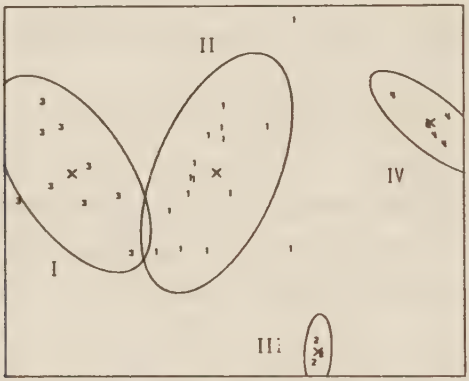

Abb. 31: Darstellung von Segmenten der nach Abb. 25 gefundenen Klassen nach kanonischer Transformation. Die 3. Komponente ist nicht gezeichnet. Die Klassen I und II sind schlecht zu trennen. Falls die Merkmale normalverteilt wären, müßten 99% aller Segmente innerhalb der Ellipsen liegen ($p(\underline{x}|k) = 0{,}01$).

Die Ellipsen als Kurven konstanter Wahrscheinlichkeitsdichte sind für $p(\underline{x}|k) = 0{,}01$ bestimmt. Obwohl die der Berechnung zugrundeliegende Annahme einer Normalverteilung wohl kaum erfüllt ist, geben die Ellipsen doch ein gutes Bild über die Größe und Lage der Klassen. Die Klassen I und II sind schlecht zu trennen, wie wir schon früher gesehen haben. Die Klassen III und IV sind erstaunlicherweise sehr gut getrennt. Wir müssen jedoch berücksichtigen, daß

die 3. Komponente, die zu dem 3. Eigenwert gehört, nicht dargestellt ist.

Wir werden im Folgenden nur noch 3-Klassen-Probleme zur Illustration verwenden, da hier die Merkmalsvektoren nach der kanonischen Transformation durch 2 Komponenten vollständig beschrieben werden können.

Klassifikatoren

Klassifikatoren dienen dazu, Entscheidungsflächen zwischen zwei Klassen k und i festzulegen. Einen einfachen linearen Klassifikator erhalten wir durch Vergleich der geometrischen Distanzen eines Merkmalvektors $\underline{x}$ zu den jeweiligen Klassenmittelpunkten $\underline{m}_i$ und $\underline{m}_k$; $\underline{x}$ wird derjenigen Klasse zugeordnet, zu der er die kleinste Distanz hat. Als Abstandsmaß d_k zum Mittelpunkt der Klasse k können wir das Quadrat des euklidschen Abstandes benutzen (Meyer-Brötz, Schürmann 1970):

$$d_k = (\underline{x} - \underline{m}_k)^T (\underline{x} - \underline{m}_k) \qquad (25)$$

Die Entscheidungsflächen sind definiert als $d_k = d_i$ oder $d_k - d_i = 0$. Wir erhalten also:

$$(\underline{x} - \underline{m}_k)^T(\underline{x} - \underline{m}_k) - (\underline{x} - \underline{m}_i)^T(\underline{x} - \underline{m}_i) = 0$$

und

$$2(\underline{m}_i - \underline{m}_k)^T \underline{x} - (\underline{m}_i - \underline{m}_k)^T(\underline{m}_i + \underline{m}_k) = 0$$

Dieses ist die Gleichung einer Ebene, die sich allgemein

darstellen läßt als:

$$\underline{w}^T\underline{x} + w_o = 0$$

Aus dem Vektor $\underline{w}$ läßt sich der Normalenvektor, $\underline{n}$, der Ebene sowie der Abstand a vom Nullpunkt bestimmen:

$$\underline{n} = \frac{\underline{w}}{|\underline{w}|} \quad \text{und} \quad a = -\frac{w_o}{|\underline{w}|}$$

Dieser lineare euklidsche Klassifikator ist in Abb. 32 dargestellt. Die verschiedenen Klassen repräsentieren EEGs von Kindern verschiedener Altersgruppen, Klasse 1 von 6-jährigen, Klasse 3 von 10-jährigen und Klasse 4 von 12-jährigen. Die Merkmalsvektoren bestehen aus den ersten

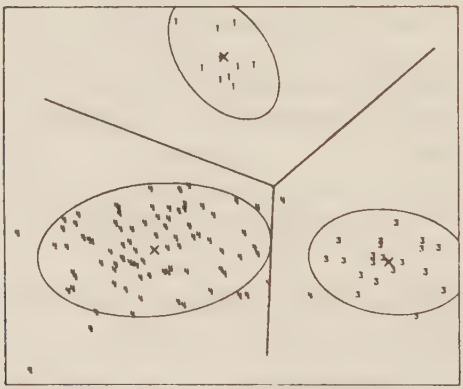

Abb. 32: Klassifikation von EEGs verschiedener Altersgruppen mit einem euklidschen Klassifikator. Klasse 1: 10 EEGs von 6-jährigen; Klasse 3: 20 EEGs von 10-jährigen; Klasse 4: 85 EEGs von 12-jährigen. Die Ellipsen als Kurven konstanter Wahrscheinlichkeitsdichte sind für $p(\underline{x}|k) = 0{,}01$ bestimmt.

zwölf Werten der Autokorrelationsfunktion. Die Trennebene steht senkrecht auf der Verbindungslinie zwischen $\underline{m}_i$ und $\underline{m}_k$, da außerdem für $\underline{x} = 1/2(\underline{m}_i+\underline{m}_k)$ die Ebenengleichung erfüllt ist, muß die Trennebene die Mittelsenkrechte zwischen beiden Klassen bilden.

Das Abstandsmaß (25) ist insofern nicht befriedigend, als große Werte der Komponenten von $\underline{x}$, die i.a. mit größeren Varianzen behaftet sind, große Beiträge zum Abstandsmaß liefern. Es empfiehlt sich daher, die einzelnen Komponenten zu wichten, damit sie, unabhängig von ihrer Größe, vergleichbare Beiträge zum Abstandsmaß beisteuern. Das Abstandsmaß wird nach Multiplikation mit einer Matrix von Gewichtsfaktoren:

$$d_k = (\underline{x} - \underline{m}_k)^T \underline{\underline{G}}_k (\underline{x} - \underline{m}_k) \qquad (26)$$

Die Elemente von $\underline{\underline{G}}_k$ werden so gewählt, daß "die zu erwartenden Abstände d_k innerhalb der Klasse k möglichst klein werden" (Meyer-Brötz, Schürmann 1970). Die Lösung dieser Minimierungsaufgabe führt zu einer Gewichtsmatrix:

$$\underline{\underline{G}}_k \sim \begin{pmatrix} 1/\sigma^2_{1k} & & 0 \\ & \ddots & \\ 0 & & 1/\sigma^2_{pk} \end{pmatrix}$$

Die Gewichtsfaktoren sind umgekehrt proportional zur Varianz der einzelnen Merkmale; Merkmale mit großer Streuung liefern nur einen kleinen Beitrag, während solche mit geringer Varianz einen großen Beitrag zum

Abstandsmaß liefern. Gleichung (26) führt auf einen quadratischen Klassifikator, der linearisiert werden kann, wenn über die Varianzen der Merkmale aller Klassen gemittelt wird:

$$\sigma_i = \frac{1}{g} \sum_{k=1}^{g} \sigma_{ik}$$

Die Gewichtsmatrix $\underline{\underline{G}}_k$ wird klassenunabhängig:

$$\underline{\underline{G}}_k = \underline{\underline{G}} = \begin{pmatrix} 1/\sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_p^2 \end{pmatrix}$$

Die Trennebene läßt sich aus $d_k - d_i = 0$ bestimmen:

$$(\underline{m}_k - \underline{m}_i)^T \; \underline{\underline{G}} \; \underline{x} = \frac{1}{2} (\underline{m}_k - \underline{m}_i)^T \; \underline{\underline{G}} \; (\underline{m}_k + \underline{m}_i) = 0$$

Merkmale mit kleiner Varianz innerhalb einer Klasse werden stärker gewichtet. Dies ist gut in Abb. 33 an der Entscheidungsgrenze zwischen den Klassen 1 und 4 im Vergleich zu Abb. 32 zu erkennen.

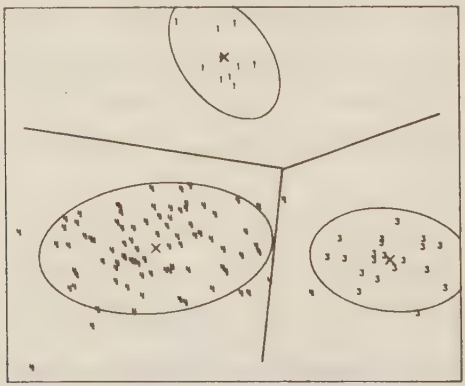

Abb. 33: Klassifikation von EEGs mit einem gewichteten euklidschen Klassifikator. EEGs wie in Abb. 32.

Der Vorzug linearer Klassifikatoren gegenüber nichtlinearen besteht in ihrer einfachen Realisierbarkeit. Nichtlineare Klassifikatoren lassen sich i.a. nur bestimmen, falls die Wahrscheinlichkeitsverteilungen der Komponenten des Merkmalvektors bekannt sind. Speziell für Normalverteilungen lassen sich die optimalen Trennflächen bestimmen (Nagy 1968), sie sind hyperquadratisch; für 2 Dimensionen sind es Kreise, Ellipsen, Parabeln oder Hyperbeln.

Zur Bestimmung der Trennflächen werden Funktionen der Form $d_k = P_k \; p(\underline{x}|k)$ oder monotone Funktionen davon bestimmt:

$$d_k = \ln(P_k \; p(\underline{x}|k)) + \frac{p}{2} \ln(2\pi)$$

Mit einer Wahrscheinlichkeitsdichte nach Gleichung (24) erhalten wir:

$$d_k = \ln P_k - \frac{1}{2} \ln|\underline{\underline{K}}_k| - \frac{1}{2}(\underline{x}-\underline{m}_k)^T \; \underline{\underline{K}}_k^{-1} (\underline{x}-\underline{m}_k)$$

Die Trennfläche ist gegeben durch $d_k - d_i = 0$:

$$\ln \frac{P_k}{P_i} - \frac{1}{2} \ln \frac{|\underline{\underline{K}}_k|}{|\underline{\underline{K}}_i|} - \frac{1}{2}(\underline{x}-\underline{m}_k)^T \; \underline{\underline{K}}_k^{-1} (\underline{x}-\underline{m}_k) + \frac{1}{2}(\underline{x}-\underline{m}_i)^T \underline{\underline{K}}_i^{-1} (\underline{x}-\underline{m}_i) = 0$$

Dieser Klassifikator enthält von den Merkmalsvektoren unabhängige, konstante Glieder, sowie lineare und quadratische Glieder. In Abb. 34 sind die quadratischen Trennlinien eingezeichnet, die, falls die Komponenten der Merkmalsvektoren normalverteilt sind, eine optimale

Klassentrennung erlauben. Im Vergleich zu den linearen Klassifikatoren liegen alle Merkmalsvektoren in den richtigen Sektoren.

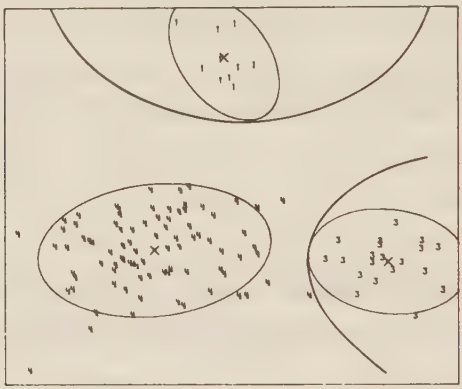

Abb. 34: Klassifikation mit einem quadratischen Klassifikator. EEGs wie in Abb. 32.

Klassifikationsfehler

Die Wahrscheinlichkeit zu Fehlklassifikationen kann zur Beurteilung von den Problemen angemessenen Klassifikatoren oder zur Merkmalsauswahl benutzt werden. Von noch größerer Wichtigkeit ist die Abschätzung des Fehlers, wenn nicht nur relative Vergleiche durchgeführt werden sollen, sondern wenn eine verläßliche Diagnose gestellt werden muß. Unglücklicherweise läßt sich die Wahrscheinlichkeit einer Fehlklassifikation nur als sehr komplexe Größe darstellen (Fukunaga 1972), sodaß der allgemeine theoretische Ausdruck bei praktischen Anwendungen keine große Hilfe bedeutet. Der Klassifikationsfehler kann daher i.a.

nur experimentell abgeschätzt werden, er hängt vom Klassifikator und von der Verteilung der Merkmalsvektoren ab.

Die einfachste Technik zur Abschätzung dieses Fehlers wird Resubstitutions- oder Reklassifikationsmethode genannt. Wir gehen davon aus, daß N Merkmalsvektoren aus g Klassen bekannt sind. Wie oben geschrieben, können Entscheidungsflächen zwischen den einzelnen Klassen bestimmt werden. Zur Fehlerabschätzung werden anschließend die N bekannten Merkmalsvektoren klassifiziert, die richtigen Zuordnungen werden aufsummiert und geben die Klassifikationsrate. In den Abbn. 32 und 33 werden jeweils 2 Vektoren aus Klasse 4 der 3. Klasse zugeordnet, die Klassifikationsrate ist kleiner als 100%, in Abb. 34 beträgt sie genau 100%.

Diese Methode führt leicht zu falschen Ergebnissen; sie liefert immer eine zu optimistische Schätzung der Fehlklassifikationswahrscheinlichkeit, besonders wenn nur wenig Merkmalsvektoren zur Berechnung der Trennflächen vorhanden sind (Lachenbruch, Mickey 1968). Diese Methode sollte daher nur für relative Vergleiche benutzt werden.

Ein anderes Verfahren teilt von Anfang an die bekannten Merkmalsvektoren in eine Lern- und eine Testgruppe auf. Nur aus der Lerngruppe werden die Trennflächen bestimmt; mit der Testgruppe wird anschließend der Klassifikations-

fehler abgeschätzt. Nach welchen Kriterien richtet sich die Größe beider Gruppen? Wenn die Lerngruppe groß ist, werden die Trennflächen gut bestimmt, dafür variiert der Klassifikationsfehler stark. Ist die Lerngruppe klein, sind die Trennflächen nur unsicher zu bestimmen. Diese Methode geht mit den vorhandenen Merkmalsvektoren sehr unökonomisch um.

Um die Merkmalsvektoren effektiver auszunutzen, können wir so vorgehen: Aus der Menge aller Vektoren nehmen wir nur einen heraus und berechnen mit den übrigen N-1 Vektoren die Trennflächen, die wir mit dem unbenutzten Vektor anschließend testen. Diese Methode heißt totale Kreuzvalidierung oder auch "leaving-one-out" (Fukunaga 1972) oder "hold-one-out" Methode (Lachenbruch, Mickey 1968). Die Möglichkeiten zur Bestimmung des Klassifikationsfehlers werden offensichtlich optimal ausgenutzt, die Trennflächen müssen allerdings N-mal bestimmt werden, was bei einem quadratischen Klassifikator jedesmal die Inversion der Kovarianzmatrix bedeutet. Wenn sich die Komponenten des Merkmalsvektors durch eine mehrdimensionale Normalverteilung beschreiben lassen, ist jedoch nur eine einzige Matrixinversion nötig (Fukunaga 1972). In diesem Fall dauert die Berechnung des Klassifikationsfehlers nicht wesentlich länger als mit der Reklassifikationsmethode.

In Abb. 35 sind beide Methoden miteinander verglichen. Zur Bestimmung der Klassifikationsraten wurden jeweils

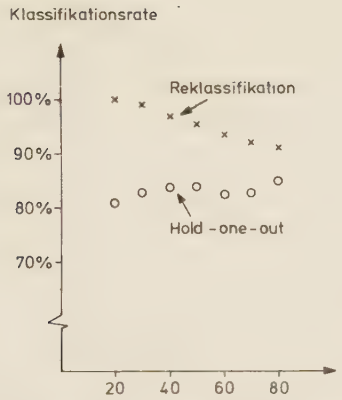

Abb. 35: Vergleich der Reklassifikations- und "hold-one-out"-Methode zur Bestimmung der Fehlklassifikationswahrscheinlichkeit. Die Klassifikationsrate gibt die Zahl der richtigen Zuordnungen an. Vier Klassen, mit je 20 bis 80 EEGs. Merkmalsatz: Prädiktor-Polynom.

vier Klassen mit einer gleich großen Anzahl von EEGs benutzt, die für jede Klasse zwischen 20 und 80 variierte. Die Klassifikation erfolgte mit quadratischen Trennflächen. Wenn die Zahl der Merkmalsvektoren groß ist, wird der Unterschied zwischen beiden Verfahren kleiner, ein Ergebnis, das auch Fukunaga (1972) angibt.

Mit der Klassifikationsrate läßt sich die Güte der Klassifikatoren abschätzen. Als Beispiel wurden 340 EEGs aus vier Klassen untersucht. Jedes EEG wird durch einen 12-dimensionalen Merkmalsvektor beschrieben. Quadratische Klassifikatoren liefern das beste Ergebnis.

Tabelle 3

Klassifikator	quadratisch	linear, gewichtet	linear
Reklass. Rate	92 %	67 %	45 %

3.3 Wahl der Merkmale

Die Wahl der Komponenten der Merkmalsvektoren hängt nicht so sehr von ihrer möglichst leichten Interpretierbarkeit ab, sondern primär von den erreichbaren Klassifikationsergebnissen. Nach Erkennung einer Segmentgrenze wird - wie oben beschrieben - die Autokorrelationsfunktion des Segmentes abgespeichert. Da wir quadratische Entscheidungsflächen nehmen wollen, weil sie die besten Klassifikationsergebnisse liefern, kann man die Autokorrelationskoeffizienten so transformieren, daß die neuen Koeffizienten näherungsweise normalverteilt sind (z-Transformation, Fisher 1956):

$$z_i = \frac{1}{2} \ln \frac{1 + r_i}{1 - r_i} = \operatorname{arth}(r_i) \quad ; \quad r_i = \frac{R_i}{R_o}$$

Ferner können wir die bei der Lösung der Normalgleichungen mit berechneten Parcor-Koeffizienten (Gleichung 10) benutzen und ebenfalls die z-Transformation auf sie anwenden. Wir können auch das Prädiktor-Polynom selbst nehmen, aus dem sich schließlich noch das Cepstrum rekursiv bestimmen läßt (Makhoul 1975):

$$c_n = a_n - \sum_{m=1}^{n-1} \frac{m}{n} c_m a_{n-m} \quad ; \quad 1 \leq n \leq p$$

Diese sechs Merkmalsgruppen werden auf ihre Klassifikationsgüte hin untersucht an Hand von 340 EEGs aus vier Klassen.

Wilks Λ-Kriterium kann benutzt werden, um die Nullhypothese der Gleichheit der Mittelwerte zweier Klassen zu testen. Tabelle 4 zeigt die Ergebnisse für die Klassen 1 und 2 bei Benutzung eines quadratischen Klassifikators. Dazu sind der nach Gleichung (22) berechnete F-Testwert und die Reklassifikationsrate mit angegeben (AKF: Autokorrelationsfunktion):

Tabelle 4

	Λ	F_{12}^{157}	Reklass. Rate
AKF	.7610	4.10	84.1
Arth (AKF)	.7687	3.94	83.9
Prädiktor	.7349	4.72	90.6
Parcor	.7433	4.52	86.5
Arth (Parcor)	.7410	4.57	87.1
Cepstrum	.7387	4.63	91.2

Um eine möglichst gute Klassentrennung zu erhalten, müssen die Werte für Λ klein werden, da $\Lambda = |\underline{\underline{W}}|/|\underline{\underline{T}}|$. Das Prädiktor-Polynom und das Cepstrum liefern die kleinsten Werte. Die Nullhypothese der Gleichheit der

Mittelwerte kann mit einer Irrtumswahrscheinlichkeit von $p \leq 0.001$ bei allen sechs Merkmalsgruppen verworfen werden. Die Irrtumswahrscheinlichkeit ist umso kleiner, je größer der F-Wert ist. Auch aus der Reklassifikationsrate ist zu sehen, daß das Prädiktor-Polynom und das Cepstrum die beste Klassentrennung ermöglichen.

In Abb. 36 ist die Reklassifikationsrate als Funktion der Länge des Merkmalvektors aufgetragen. Es wurden wieder vier Klassen mit 340 EEGs und ein quadratischer Klassifikator

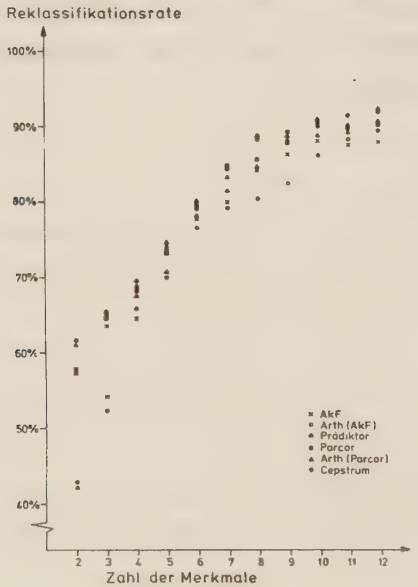

Abb. 36: Bestimmung der zur Klassifikation geeignetsten Merkmale. 340 EEGs aus vier Klassen. Für $p \geq 8$ liefern das Prädiktor-Polynom und das Cepstrum die besten Reklassifikationsraten. Bei mehr als 8 Merkmalen verändert sich die Reklassifikationsrate nur noch unwesentlich.

benutzt. Während bei einer nur kleinen Zahl der Merkmale das Cepstrum und das Prädiktor-Polynom schlechte Reklassifikationsraten liefern, sind sie für größere Merkmalszahlen ($p \geq 8$) am besten geeignet. Wir sehen außerdem, daß sich für $p \geq 8$ die Reklassifikationsrate nur noch unwesentlich ändert. Das deutet darauf hin, daß zur Beschreibung von EEGs ein Merkmalvektor mit $p = 8$ Koeffizienten ausreicht, ein Ergebnis, das wir mit völlig anderen Methoden schon früher erhalten haben.

Qualitativ kann die Klassifikationsgüte nach der kanonischen Transformation wie in Abb. 37 und Abb. 38 dargestellt werden. Die gesamte Information der drei gezeigten Altersklassen mit je 85 EEGs kann 2-dimensional gezeichnet werden. In Abb. 37 bestehen die Merkmalvektoren aus den Autokorrelationskoeffizienten, in Abb. 38 aus dem Prädiktor-

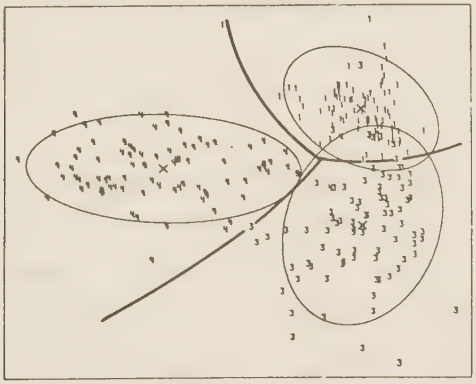

Abb. 37: Klassifikation mit einem quadratischen Klassifikator. Drei Klassen mit je 85 EEGs. $p(\underline{x}|k) = 0{,}01$. Merkmalsatz: Autokorrelationskoeffizienten.

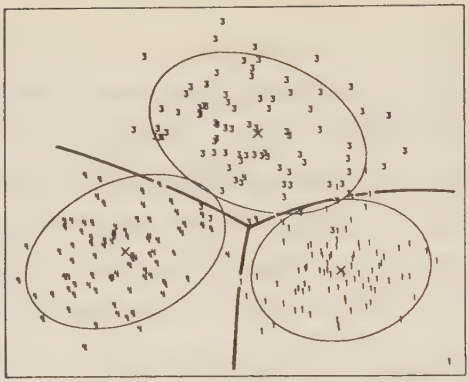

Abb. 38: Klassifikation mit einem quadratischen Klassifikator. EEGs wie in Abb. 37. Merkmalsatz: Prädiktor-Polynom. Das Reklassifikationsergebnis ist besser als das nach Abb. 37.

Polynom. Auch hier liefern das Prädiktor-Polynom und das Cepstrum die besten Ergebnisse. Quadratische Trennfunktionen sind eingezeichnet. Die Abbildungen sind von klinischen Gesichtspunkten her interessant, da sie z.B. zeigen, daß EEGs von 10-jährigen (Klasse 3) deutlich unterschieden werden können von denen von 12-jährigen Kindern (Klasse 4).

In den Abbn. 39 und 40 ist eine letzte Möglichkeit zur Gütebestimmung der Merkmale zu sehen. Die bei hierarchischer Fusion von Segmenten entstehenden Dendrogramme lassen die mit verschiedenen Merkmalen erreichbare Homogenität der Gruppierungen erkennen. Nach den Autokorrelationskoeffizienten (Abb. 39) ist schwer zu entscheiden, welche Gruppierung endgültig ausgewählt werden soll, da die Änderungen des Homogenitätsmaßes (Ordinate) immer ziemlich gleich sind;

Abb. 39: Automatische Klassifikation von Segmenten. Merkmalsatz: Autokorrelationskoeffizienten. Eine Gruppierung ist schwer festzulegen.

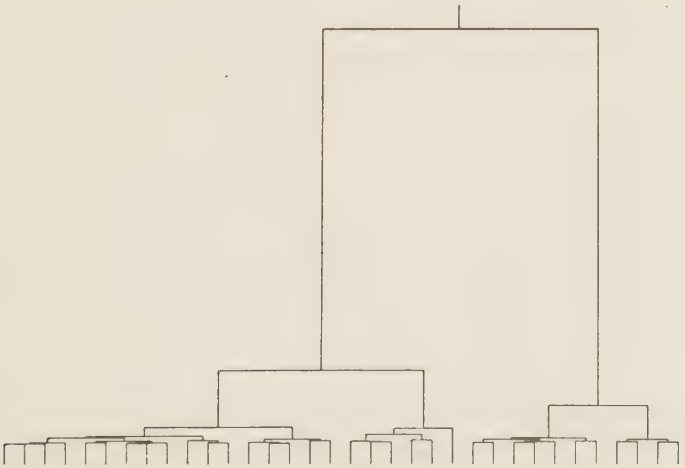

Abb. 40: Automatische Klassifikation von Segmenten. Dieselben Segmente wie in Abb. 39. Merkmalsatz: Cepstrum. Das Dendrogramm legt nahe, eine endgültige Gruppierung aus zwei Klassen zu bilden.

das Cepstrum dagegen liefert eine klare Gruppierung derselben Segmente (Abb. 40).

Von den sechs untersuchten Merkmalsgruppen haben sich das Prädiktor-Polynom und das Cepstrum als für Klassifikationsverfahren am besten geeignet herausgestellt.

4. Programme

Zur Signalanalyse, Merkmalsgewinnung und Weiterverarbeitung wurden zahlreiche Programme auf der UNIVAC 1108 der GWDG entwickelt. Alle Programme sind ausnahmslos in FORTRAN V geschrieben und benutzen nur in relativ geringem Umfang maschinenabhängige Befehle. Die untersuchten EEGs wurden auf einem PDP-12 System mit einer Abtastrate von 100 Hz digitalisiert, auf der UNIVAC in 36-Bit Worte umgewandelt und auf Magnetbändern gespeichert.

Vor jeglicher Art der Weiterverarbeitung werden die EEGs tiefpaßgefiltert mit einer oberen Frequenz von 50 Hz und neu abgetastet.

Nur das Programmsystem zur Merkmalsgewinnung wird etwas ausführlicher beschrieben. Von dem zuerst entwickelten Programmsystem zur Signalanalyse sollen nur die Optionen erwähnt werden:
Das zu untersuchende Signal kann gefiltert werden, wobei obere und untere Grenzfrequenz frei wählbar sind; es kann differenziert werden; es kann zur Berechnung der Autokorrelationsfunktion mit fünf verschiedenen Fenstern multipliziert werden. Aus einem Segment beliebiger Länge kann mit der Spline-Interpolation (Sauer, Szabó 1968) ein FFT-Spektrum berechnet werden; aus einem beliebigen Segment kann ein Prädiktor-Polynom beliebiger Länge berechnet werden, mit diesem der Prädiktionsfehler samt dessen Spektrum und Autokorrelationsfunktion. Aus dem

Prädiktor-Polynom kann das Spektrum nach Gleichung (11) und die zugehörige Autokorrelationsfunktion bestimmt werden. Schließlich ist es möglich, die Transientenerkennung mit variablen Schwellwerten zu untersuchen und Transiente zu rekonstruieren. Der Erwartungswert des quadrierten Prädiktionsfehlers (Gleichung (9)), dessen Abhängigkeit von der Länge des Prädiktor-Polynoms und die Größen AIC (Gleichung 12) und CAT (Gleichung 13) können bestimmt werden. Aus dem mit diesem Programm gewonnenen Erfahrungen wurde das Programmsystem zur Merkmalsgewinnung entwickelt.

4.1 Datenhaltungssystem

Von der Eingabedatei 'EEG' werden Original-EEGs oder Pulsationen gelesen. Wenn mehrere gleichzeitige Ableitungen vorhanden sind, werden die Daten in Blöcken zu je 512 Werten kanalweise hintereinander abgespeichert. Um Verwechselungen zu vermeiden, steht am Dateianfang ein Codewort.

Die aus dem Signal gewonnenen Merkmalsätze werden auf der Datei 'FEA' (feature) geschrieben. Für jede untersuchte Zeitreihe gibt es auf Datei 'INFO' einen Informationsblock von 100 Worten Länge, in dem Anfangs- und Endadresse der Merkmalsätze auf Datei 'FEA', die verschiedenen zur Analyse gebrauchten Parameter sowie Datum und Zeit gespeichert sind. Wird z.B. die 27. Analyse durch-

geführt, liest das Programm den Informationsblock der 26. Analyse und erhält daraus die Adresse des letzten auf Datei 'FEA' geschriebenen Merkmalsatzes.

Nachdem alle Segmente und Transiente erkannt und auf 'FEA' gespeichert sind, wird der Informationsblock der 27. Analyse auf 'INFO' geschrieben. Die Abspeicherung von Transienten und Segmenten ist in Tabelle 5 dargestellt. Von jedem Segment werden 11 Werte der Autokorrelationsfunktion gespeichert, sowie als letzter Wert die Länge des Segments, zu der die mit 10 000 multiplizierte Ableitungsnummer addiert wird. Vor jedem Segment werden eventuell vorhandene Transienten abgespeichert. Die ersten beiden Zahlen geben Anfang und Ende innerhalb des Segmentes an. Die erste Zahl wird mit der Ableitungsnummer codiert und mit einem negativen Vorzeichen zur Kennzeichnung als Transient versehen. Von der dritten Zahl an stehen die Werte des Prädiktionsfehlers selbst. Falls die Länge des Merkmalsatzes zur Abspeicherung von Transienten nicht ausreicht, wird die zweite Zahl ebenfalls mit einem negativen Vorzeichen versehen und der nächste Merkmalsatz mitbenutzt. Diese Möglichkeit wird jedoch fast nie gebraucht. Die Erfahrung hat gezeigt, daß Transienten in der Regel durch drei bis vier Werte des Prädiktionsfehlers dargestellt werden können. Eine Verwechselung von Segment und Transient ist nach diesem System ausgeschlossen, da die erste Zahl eines Segments immer positiv ist (Leistung des Segments), während sie bei Transienten immer negativ ist.

Tabelle 5

Abspeicherung von Segmenten und Transienten

Länge des Merkmalsatzes p = 12

Bedeutung \ Wort Nr.		1	2	3	4	5	6	7	8	9	10	11	12
Kanal 1	Transient	-10048	51	S_1	S_2	S_3	S_4	0	0	0	0	0	0
	Segment	R_0	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	10340
Kanal 2	Segment	R_0	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	20340
Kanal 3	Transient	-10050	-59	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
	Transient	-10060	62	S_{11}	S_{12}	S_{13}	0	0	0	0	0	0	0
	Segment	R_0	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	30340

4.2 Programme zur Merkmalsgewinnung

Ein Flußdiagramm des Programmsystems zur Erkennung und Abspeicherung von Segmenten und Transienten ist in Abb. 41 gezeigt. Nach dem Eingabedialog, der von einem Terminal geführt werden kann, wird ein Block mit 512 Werten eingelesen. Aus den ersten 2 sec des Signals (vgl. Abb. 14) wird ein Prädiktor-Polynom der Länge p = 8 und der entsprechende Prädiktionsfehler bestimmt, dessen Spitzen abgeschnitten werden, falls sie eine bestimmte Schwelle überschreiten. Aus dem geclippten Prädiktionsfehler wird die Autokorrelationsfunktion des festen Fensters und daraus das spektrale Fehlermaß für den ersten Wert berechnet. Dieser Prozeß wird parallel für alle Ableitungen durchgeführt. Der spektrale Fehler aller Ableitungen wird aufsummiert und mit einer Schwelle verglichen. Ist er kleiner, wird aus dem um einen Punkt verschobenen laufenden Fenster eine neue Autokorrelationsfunktion aus dem geclippten Prädiktionsfehler berechnet und ein neuer Wert des spektralen Fehlermaßes bestimmt. Ist dieser größer als die Schwelle, wird eine Grenze erkannt; daraufhin wird der nicht geclippte Prädiktionsfehler auf Transienten untersucht, die nach Tabelle 5 auf Datei 'FEA' geschrieben werden. Schließlich wird noch die Autokorrelationsfunktion des Segmentes für elf Werte berechnet und mit der der Segmentlänge auf Datei 'FEA' geschrieben. Falls noch genug Daten auf Datei 'EEG' vorhanden sind, beginnt der ganze Prozeß von vorn. Es ist möglich, daß sich während eines Blockes von 512 Werten das Spektrum des Signals

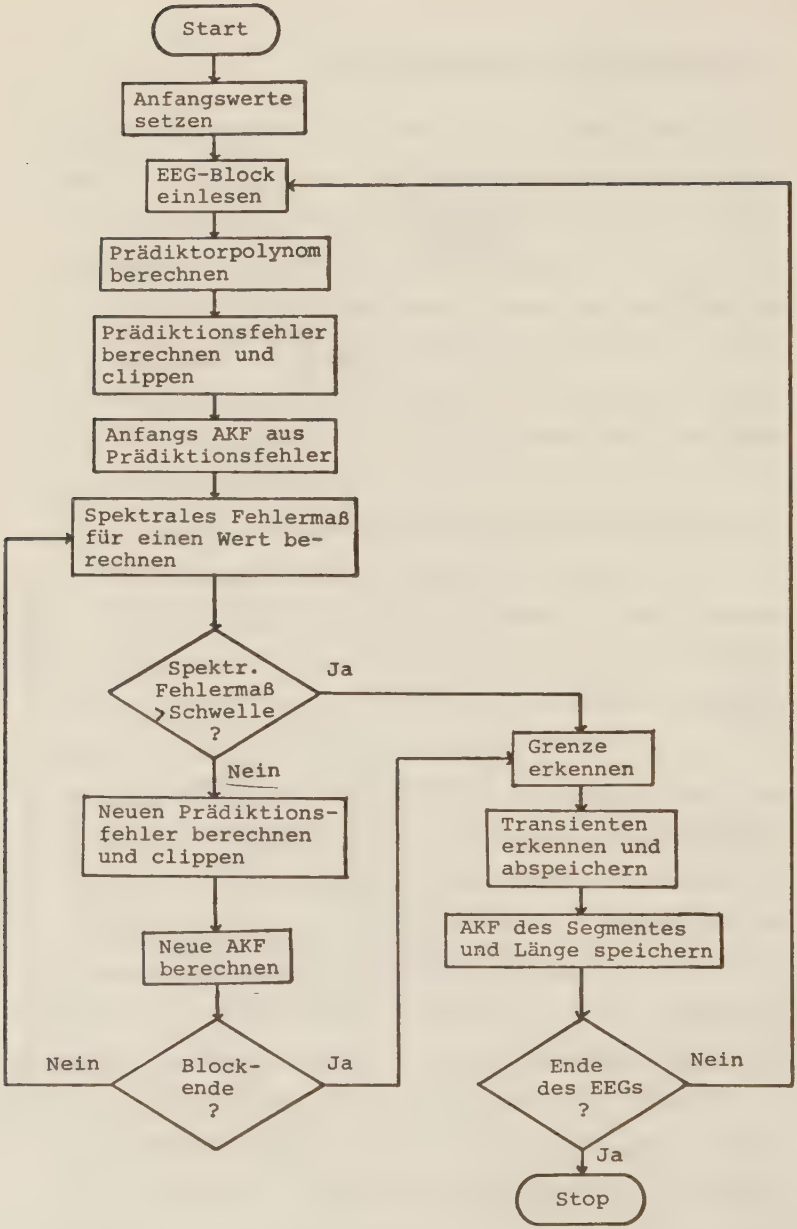

Abb. 41: Flußdiagramm des Programmsystems zur Merkmalsgewinnung.

nicht ändert; in diesem Fall wird bei Erreichen der Blocklänge eine Grenze erkannt (vgl. 1. Segmentgrenze in Abb. 14).

Obwohl dieser ganze Ablauf relativ kompliziert ist, läßt er sich für mehrere Kanäle in Echtzeit ausführen. Für ein EEG mit acht Ableitungen - ein Ausschnitt ist in Abb. 19 gezeigt - werden bei einer Dauer von 80 sec nur etwa 40 sec zur Merkmalsgewinnung und Abspeicherung benötigt. Bei normalen EEGs mit nur wenigen Segmenten und fast keinen Transienten sieht dieses Verhältnis noch sehr viel günstiger aus. Zur Untersuchung der Merkmalsätze wurden Programme entwickelt, die in das Original-Signal die erkannten Segmente und Transiente einzeichnen (vgl. Abbn.19 und 20) sowie das Signal mit den Transienten rekonstruieren können (vgl. Abbn. 21 und 22).

4.3 Programme zur Klassifikation

Für die automatische Klassifikation steht bei der GWDG ein Programmsystem zur Verfügung, das D. Wishart von der St. Andrew's Universität, Schottland, 1970 unter dem Namen CLUSTAN herausgegeben hat. Dieses Programm ist wegen zahlreicher Optionen umständlich zu benutzen und braucht viel Speicherplatz und Rechenzeit. Dafür stehen mehrere Verfahren zur Verfügung, von denen das hierarchische Verfahren von Ward (1963) gute Ergebnisse liefert.

Da die automatischen Klassifikationsverfahren unabhängig

von a-priori definierten Klassen sind, wäre es schön, wenn dieses oder ein ähnliches Verfahren mit in das System zur Merkmalsgewinnung integriert werden könnte, besonders weil die Datenreduktion erheblich zunähme. Leider müssen bei fast allen Verfahren vor dem ersten Fusionsschritt alle Merkmalsätze bekannt sein, sodaß Merkmalsgewinnung und automatische Klassifikation in Echtzeit nicht möglich sind.

Die Programme zur Diskriminanzanalyse, nämlich Berechnung und Zeichnung der Klassifikatoren, kanonische Transformation, Bestimmung der Fehlklassifikationsraten, wurden alle selbst entwickelt; da diese Programme speziell auf die Aufgabenstellung zugeschnitten sind, benötigen sie nur wenig Speicherplatz und sind sehr schnell, z.B. werden zur Bestimmung einer quadratischen Diskriminanzfunktion sowie zur anschließenden Reklassifikation von 340 Merkmalsätzen mit je 12 Merkmalen nur etwa 6 sec Rechenzeit benötigt. Eine Echtzeitklassifikation könnte nach der kanonischen Transformation durchgeführt werden, wenn vorher in der Lernphase die Trennflächen bestimmt wurden.

5. Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit Problemen der Mustererkennung von stochastischen Signalen. Als Beispiele werden hauptsächlich bioelektrische Signale und zwar Elektroenzephalogramme (EEG) benutzt. Die entwickelten Verfahren lassen sich auch auf andere stochastische Signale anwenden, was an geomagnetischen Pulsationen demonstriert wird.

Es wird angenommen, daß sich die zu untersuchenden Zeitreihen aus zwei Arten von Mustern zusammensetzen lassen, aus

a) Segmenten variabler Länge, deren quadratisches Spektrum (Power-Spektrum) sich nicht oder nur wenig ändert und
b) kurzzeitigen Veränderungen (Transienten), die einer Grundaktivität überlagert sein können.

Das vorgestellte Analysesystem kann diese Mustertypen erkennen und durch Merkmale beschreiben, die für Klassifikationsaufgaben geeignet sind.

Merkmalsgewinnung

Unter Zugrundelegung eines autoregressiven Modells kann der mit diesem Modell nicht vorhersagbare Prädiktionsfehler bestimmt werden. Aus der Autokorrelationsfunktion des Prädiktionsfehlers berechnen wir ein Maß zur Erkennung von Segmentgrenzen, aus der Größe der Werte des Prädiktionsfehlers selbst kann auf kurzzeitige Veränderungen geschlossen werden. Die Merkmalsgewinnung ist für EEGs (20msec Abtast-

intervall) für maximal 16 gleichzeitige Ableitungen durchführbar. Die Datenreduktion hängt direkt mit dem Informationsgehalt des Signals zusammen: werden viele Segmente und Transiente erkannt - ist also die Information groß - ist die Datenreduktion gering. Ein typischer Wert der Datenreduktion ist 50:1, wenn keine Transiente vorhanden sind.

Klassifizierung

Durch eine automatische Klassifikation (Clusteranalyse) der aus den Segmenten gewonnenen Merkmale ist es möglich, die Datenreduktion um mindestens den Faktor 10 zu steigern. Eine Methode zur hierarchischen Gruppierung liefert gute Ergebnisse.

Um die Merkmale a-priori definierten Klassen zuzuordnen, werden Methoden der Diskriminanzanalyse benutzt. Quadratische Klassifikatoren zeigen die besten Ergebnisse.

Die Auswahl der Merkmale hat mehr nach den erzielbaren Klassifikationsergebnissen als nach ihrer Interpretierbarkeit zu erfolgen. Als am besten zur Klassifikation von Segmenten geeignet haben sich das Prädiktor-Polynom und das Cepstrum erwiesen.

LITERATURVERZEICHNIS

Akaike, H., Autoregressive model fitting for control, Ann. Inst. Statist. Math., 23, 163-180, 1971.

Atal, B.S., and M.R. Schroeder, Adaptive predictive coding of speech signals, Bell Syst. Techn. J., 49, 1973-1968, 1970.

Berger, H., Über das Elektrenkephalogramm des Menschen, Archiv für Psychiatrie und Nervenkrankheiten, 87, 527-570, 1929.

Bock, H.H., Automatische Klassifikation, Vandenhoek und Ruprecht, Göttingen, 1974.

Burg, J.B., Maximum entropy spectral analysis, 37th Meeting of the Society of Exploration Geophysicists, Oklahoma, 1967.

Burg, J.B., A new analysis technique for time series data, NATO Advanced Study Institute on Signal Processing, 1968.

Cavazza, B., F. Ferrillo, G. Rosadini, and W. Sannita, A computer analysis of human interictal epileptic spikes, in: Epilepsy (Hrsg. P. Harris et al.), 65-73, Livingstone, Edinburgh, 1974.

Chen, W.Y., and G.R. Stegen, Experiments with maximum entropy power spectra of sinusoids, J. Geophys. Res., 79, 3019-3022, 1974.

Cobb, W.A., The normal adult EEG, in: Electroencephalography, (Hrsg. D. Hill et al.), 232-249, MacMillan, New York, 1963.

Cooley, W.W., and P.R. Lohnes, Multivariate data analysis, Wiley and Sons, New York, 1971.

Creutzfeldt, O.D., J.M. Fuster, H.D. Lux und A. Nacimiento, Experimenteller Nachweis von Beziehungen zwischen EEG-Wellen und Aktivität cortikaler Nervenzellen, Naturwissenschaften, 51, 166-167, 1964.

Drägert, H., Maximum entropy analysis for geomagnetic variations, Kolloquium Erdmagnetische Tiefensondierung, (Hrsg. A. Berkthold), Grafrath, 1974.

Dubes, R.C., Pattern recognition with clinical electroencephalograms, Div. of Eng. Research, Michigan State University, AFOSR-1023-67B, 1970.

Duquesnoy, A.J., Analysis of non-stationary electroencephalograms, using a modified Kalman filter, Institute of Medical Physics, TNO, Utrecht, 1975.

Fenwick, P.B.C., P. Michie, J. Dollimore, and G.W. Fenton, Mathematical simulation of the electroencephalogram using an autoregressive series, Int. J. Bio. Comp., 2, 281-307, 1971.

Fisher, R.A., Statistische Methoden für die Wissenschaft, Oliver and Boyd, London, 1956.

Flik, T., Verfahren zur automatischen Erkennung von Mustern in biomedizinischen Signalen mit spezieller Anwendung auf Spike-Wave-Komplex-Erkennung im Elektroenzephalogramm, Diss., Fachbereich Kybernetik, TU Berlin, 1973.

Fukunaga, K., Introduction to statistical pattern recognition, Academic Press, New York, 1972.

Gersch, W., Spectral analysis of EEG's by autoregressive decomposition of time series, Math. Biosciences, 7, 205-222, 1970.

Isaksson, A., On time variable properties of EEG signals examined by means of a Kalman filter method, Royal Inst. Tech., Technical report No. 95, Stockholm, 1975a.

Isaksson, A., Examples of EEG signals with time-varying spectra analysed by means of a Kalman filter method, Royal Inst. Tech., Technical report No. 96, Stockholm, 1975b.

Jackson, P.L., Truncations and phase relationships of sinusoids, J. Geophys. Res., 72, 1400-1403, 1967.

Jenkins, G.M., and D.G. Watts, Spectral analysis and its applications, Holden Day, San Francisco, 1969.

Kalman, R.E., A new approach to linear filtering and prediction problems, Trans. ASME, J. Basic Eng., 82, 35-45, 1960.

Kromer, R.E., Asymptotic properties of the autoregressive spectral estimator, Diss., Dept. of Statistics, Stanford Univ., 1969.

Lachenbruch, P.A., and R.A. Mickey, Estimation of error rates in discriminant analysis, Technometrics, 10, 1-11, 1968.

Levinson, N., The Wiener rms (root mean square) error criterion in filter design and prediction, in: N. Wiener, Extrapolation, interpolation, and smoothing of stationary time series, Wiley and Sons, New York, 1949.

Lopes da Silva, F.H., A. Dijk, and H. Smits, Detection of nonstationarities in EEG's using the autoregressive model - an application to EEG's of epileptics, in: CEAN: Computerized EEG Analysis, (Hrsg. G. Dolce et al.), 180-199, Fischer, Stuttgart, 1975.

Makhoul, J., Linear prediction: a tutorial review, Proc. IEEE, 63, 561-580, 1975.

Makhoul, J., and J.J. Wolf, Linear prediction and spectral analysis of speech, Bolt Beranek and Newman, Report No. 2304, Cambridge, Mass., 1972.

Markel, J.D., A.H. Gray, and H. Wakita, Linear prediction of speech - theory and practice, Speech Comm. Res. Lab., SCRL Monograph No. 10, Santa Barbara, 1973.

Mathieu, M., Analyse de l'électro-encéphalogramme par modélisation prédictive, École Nationale Supérieure des Télécommunications, ENST, Paris, 1974.

Mathieu, M., W. Tirsch, and S.J. Pöppl, Multichannel online EEG analysis by means of an autoregressive model, with applications, Symposion: The Quantification of the Electroencephalogram II, Konstanz, 1976 (im Druck).

Matoušek, M., Frequency and correlation analysis, in: Handbook of Electroencephalography and Clinical Neurology, (Hrsg. M.A.B. Brazier et al.), Bd. 5A, Elsevier, Amsterdam, 1973.

Matoušek, M., I. Petersen, and S. Friberg, Automatic assessment of randomly selected routine EEG records, in: CEAN: Computerized EEG Analysis, (Hrsg. G. Dolce et al.), 424-431, Fischer, Stuttgart, 1975.

Meyer-Brötz, G. und J. Schürmann, Methoden der automatischen Zeichenerkennung, Oldenbourg,München, 1970.

Nagy, G., State of the art in pattern recognition. Proc. IEEE, 56, 836-862, 1968.

Parzen, E., Some recent advances in time series modeling, IEEE Trans. Auto. Cont., AC-19, 723-730, 1974.

Rao, C.R., Lineare statistische Methoden und ihre Anwendungen, Akademie Verlag, Berlin, 1973.

Rappelsberger, P., and H. Petsche, Spectral analysis of the EEG by means of autoregression, in: CEAN: Computerized EEG Analysis, (Hrsg. G. Dolce et al.), 27-40, Fischer, Stuttgart, 1975.

Robinson, E.A., Predictive decomposition of time series with application to seismic exploration, Geophysics, 32, 418-484, 1967.

Robinson, E.A., and S. Treitel, Principles of digital Wiener filtering, Geophysical Prospecting, 15, 311-333, 1969.

Sauer, R. und I. Szabó, Mathematische Hilfsmittel des Ingenieurs, Teil III, in: Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, 141, Springer, Berlin, 1968.

Smylie, D.E., G.K.C. Clarke, and T.J. Ulrych, Analysis of irregularities in the Earth's rotation, in: Methods in Computational Physics, (Hrsg. B.A. Bolt), 13, 391-430, 1973.

Tretter, S.A., and K. Steiglitz, Power-spectrum identification in terms of rational models, IEEE Trans. Auto. Cont., AC-12, 185-188, 1967.

Ward, J.H., Hierarchical grouping to optimize an objective function, Amer. Statist. Assoc., 58, 236-244, 1963.

Wilks, S.S., Mathematical statistics, Wiley and Sons, New York, 1962.

Wishart, D., Mode analysis: a generalization of nearest neighbour which reduces chaining effects, in: Numerical Taxonomy, (Hrsg. A.J. Cole), 282-311, Academic Press, London, 1969.

Zetterberg, L.H., Experience with analysis and simulation of EEG signals with parametric description of spectra, in: Automation of Clinical Electroencephalography, (Hrsg. P. Kellaway et al.), 161-201, North-Holland, Amsterdam, 1973.

Herrn Prof. Dr. M. Siebert danke ich für die Themenstellung zu dieser Arbeit und sein Interesse am Fortgang der Untersuchungen.

Herrn Prof. Dr. O.D. Creutzfeldt danke ich für die Möglichkeit, die Untersuchungen am Max-Planck-Institut für Biophysikalische Chemie, Abteilung Neurobiologie, ausführen zu können, sowie für die stetige Förderung in zahlreichen anregenden Gesprächen.

Herrn G. Bodenstein danke ich ganz besonders für die häufigen Diskussionen und zahlreichen Anregungen, mit denen er zum Gelingen der Arbeit beigetragen hat.

Herrn Dr. H.W. Strube danke ich für zahlreiche Diskussionen.

Mein Dank für die Überlassung der gezeigten Signale gilt Herrn Dr. H.G. Lenard (Kinderklinik Göttingen), Herrn H.A. Oldenbürger, Herrn P.M. Arnold (Psychologisches Institut Göttingen), Herrn Dr. C. Benninger (Kinderklinik Heidelberg) sowie Herrn Dr. O. Hillebrand (Geophysikalisches Institut).

Hans Michael Praetorius

Lebenslauf

7. Juni 1946: in Goslar geboren.

1953 - 1956: Besuch der Schillerschule in Goslar.

1956 - 1966: Besuch des Ratsgymnasiums in Goslar, dort Abitur.

1966 - 1972: Studium der Physik an der Technischen Universität Braunschweig.
Vordiplom: 1969
Diplom: 1972.

1972 - 1974: Wissenschaftlicher Angestellter an der Neurochirurgischen Universitätsklinik Düsseldorf in der Fachgruppe: Einsatz der EDV in der Schwerkrankenüberwachung.

seit 1974: Stipendiat am Max-Planck-Institut für Biophysikalische Chemie, Abteilung Neurobiologie, Göttingen - Nikolausberg.